



Computing as a Tool for High Energy and Nuclear Physics

Stephen Wolbers

Fermilab

April 25, 2001



Outline

- Introduction to High Energy and Nuclear Physics (HENP) and HENP Computing
 - Accelerators, detectors, data volumes
- Data Rates, Facilities, Techniques
 - Run 2a at Fermilab
 - RHI C/JLAB Experiment Offline Computing
 - Lattice Gauge Calculations
 - Accelerator Design
- Future
 - Run 2b at Fermilab
 - Data GRIDs and Distributed Computing



In Silica Fertilization

All Science Is Computer Science

By GEORGE JOHNSON

EXCEPT for the fact that everything, including DNA and proteins, is made from quarks, particle physics and biology don't seem to have a lot in common. One science uses mammoth particle accelerators to explore the subatomic world; the other uses petri dishes, centrifuges and other laboratory paraphernalia to study the chemistry of life. But there is one tool both have come to find indispensable: supercomputers powerful enough to sift through piles of data that would crush the unaided mind.

Last month both physicists and biologists made announcements that challenged the tenets of their fields. Though different in every other way, both discoveries relied on the kind of intense computer power that would have been impossible to marshal just a few years ago. In fact, as research on so many fronts is becoming increasingly dependent on computation, all science, it seems, is becoming computer science.

"Physics is almost entirely computational now," said Thomas B. Kepler, vice president for academic affairs at the Santa Fe Institute, a multidisciplinary research center in New Mexico. "Nobody would dream of doing these big accelerator experiments without a tremendous amount of computer power to analyze the data."

New York Times,
Sunday, March 25, 2001

April 25, 2001

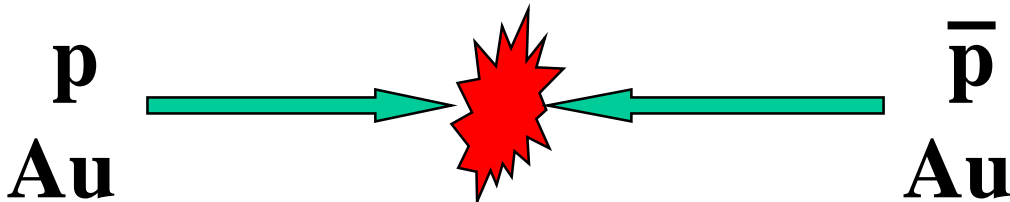
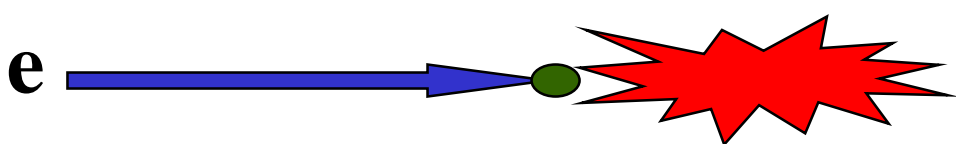


Introduction

- Particle and nuclear physics use beams of particles striking targets to study the fundamental nature of matter and interactions.
- Advances in the field have come from:
 - Higher Energy Particles and Interactions
 - More collisions per unit time and space (luminosity)
 - Better detectors
 - More sensitivity, more granular, fewer cracks, lower deadtime, more radiation-hard
 - More “events” saved to storage (disk or tape)
 - More sophisticated analysis of “events”
 - Better simulation of the beams, collisions, and detector
 - Advances in Theory

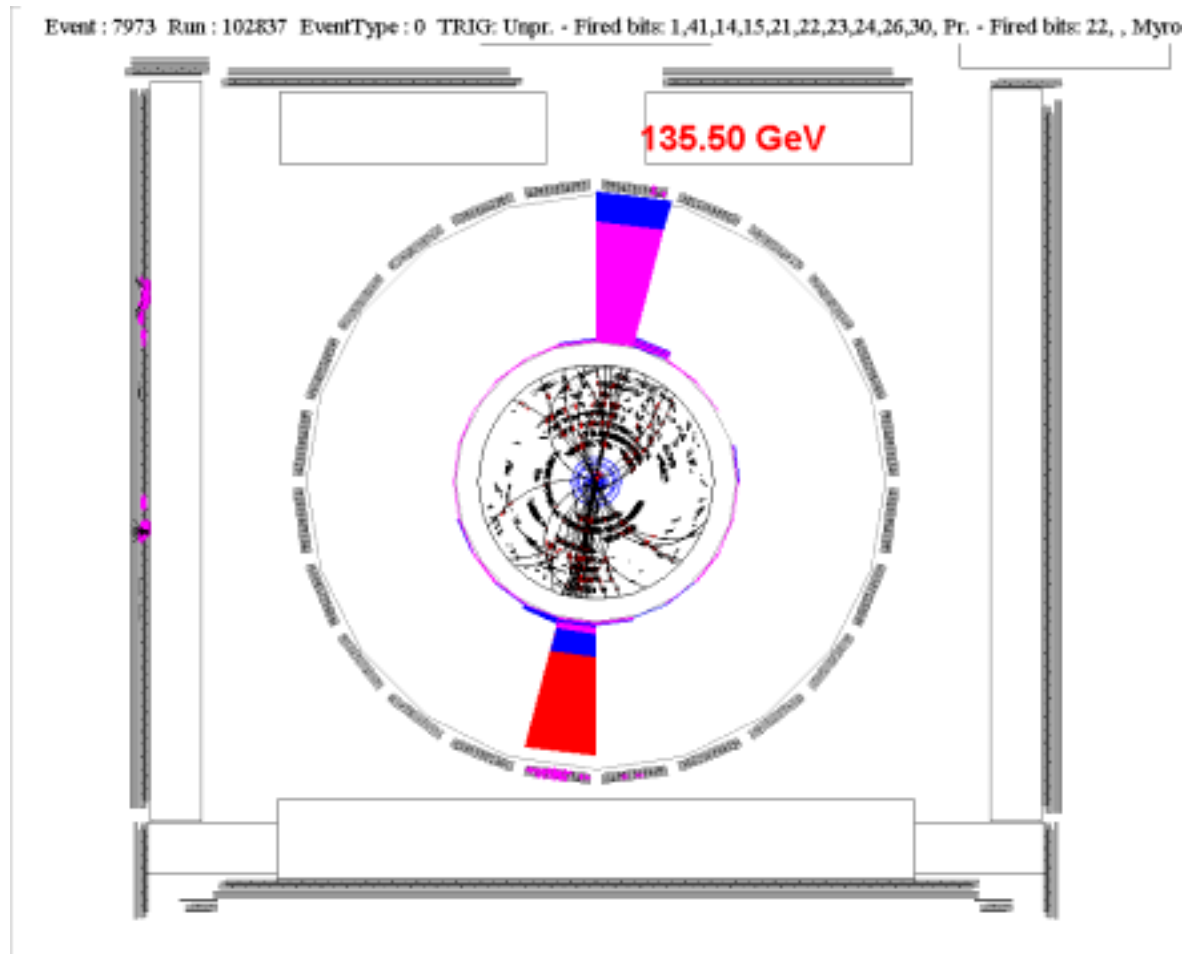


Collisions Simplified

- Collider: 
- Fixed-Target: 

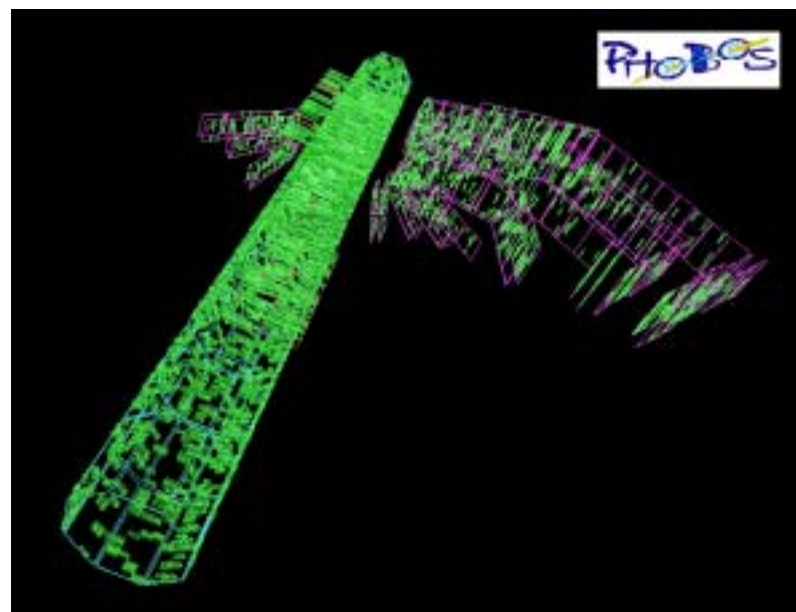
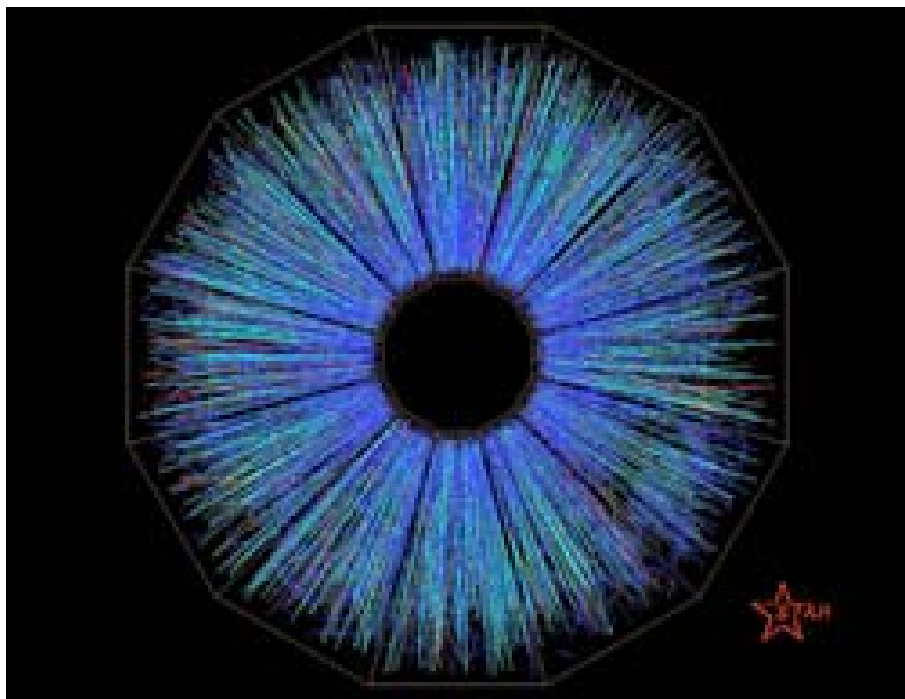


Collisions – CDF/Fall 2000





Collision at RHIC from STAR and PHOBOS





Particle Acceleration

- Particle acceleration occurs in a multi-step process
- Example: Fermilab
 - Cockcroft-Walton
 - Linac
 - Booster
 - Main Injector
 - Tevatron
- Anti-Protons
 - Accumulator/debuncher
 - Recycler



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium



Cockcroft-Walton



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

10



Linac



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

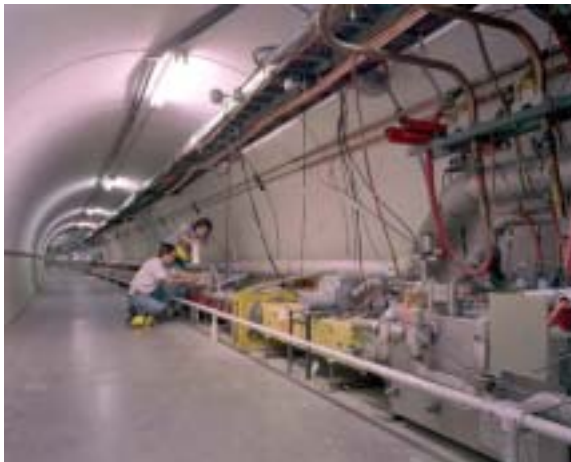
11



Booster



Main Injector



Tevatron



Antiproton Source



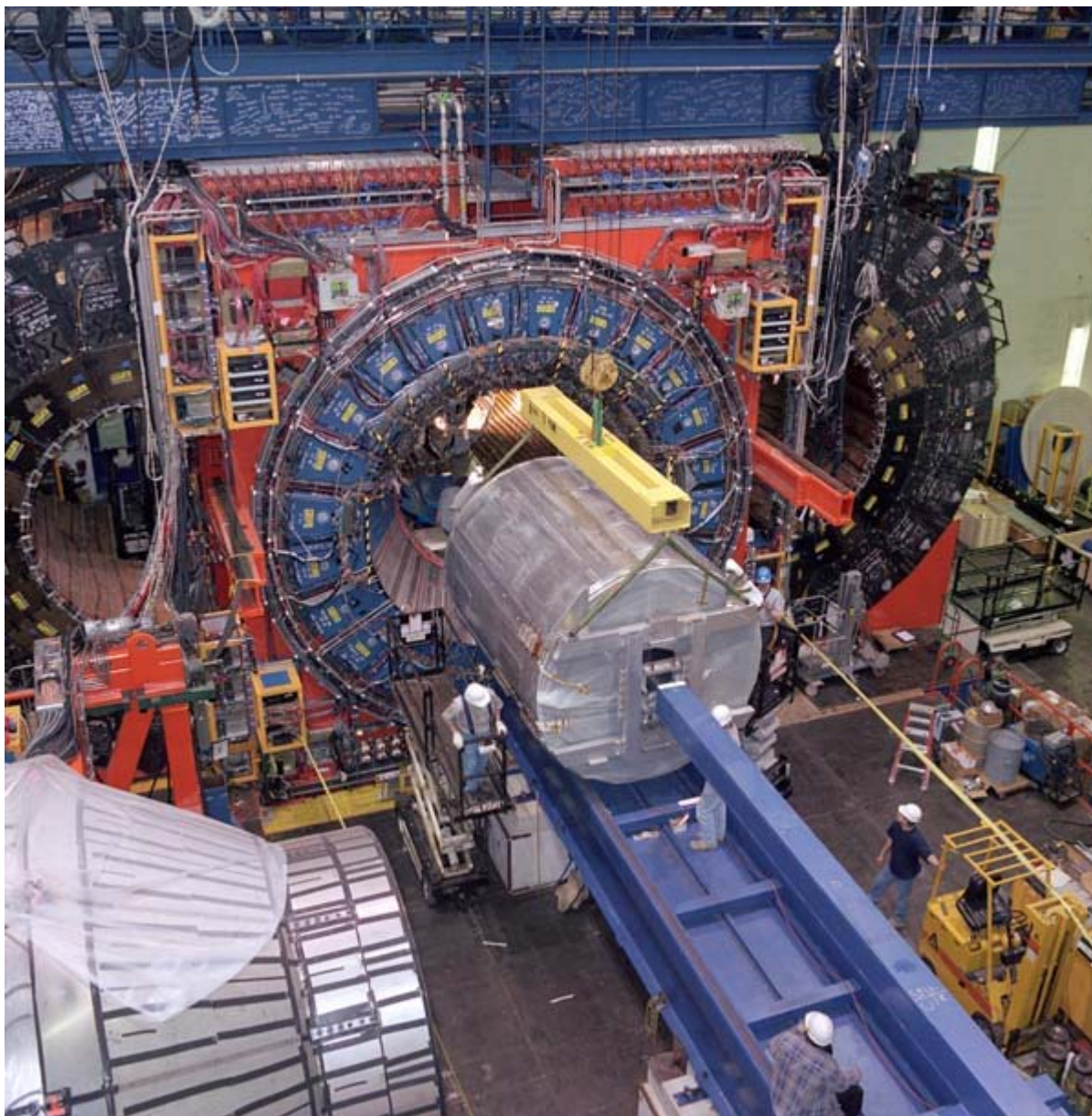
Recycler





Particle Detectors

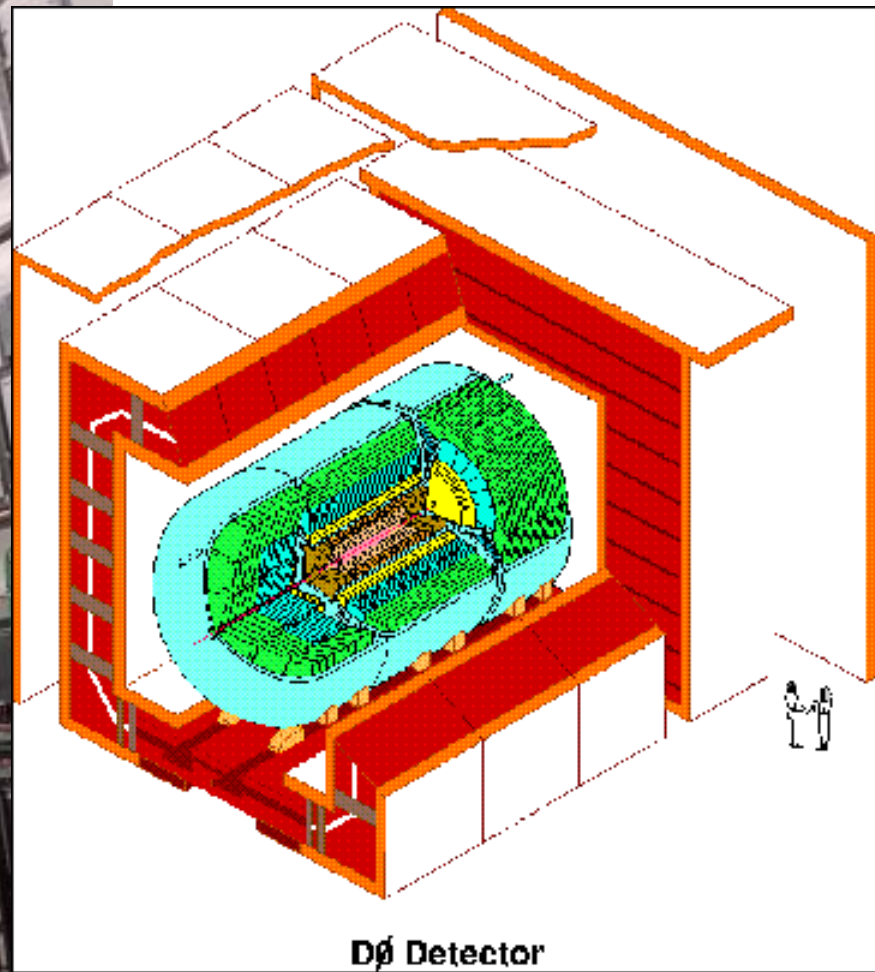
- Large detectors are used to “see” the interactions of particles.
- Consist of subdetectors which record information about particle position, energy and momentum.
 - Used to measure the number and types of particles coming from collisions.
 - Also used to identify particle decays.
 - $e, \mu, \pi, \gamma, p, K, W, Z, b, c, \dots$



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

15



April 25, 2001

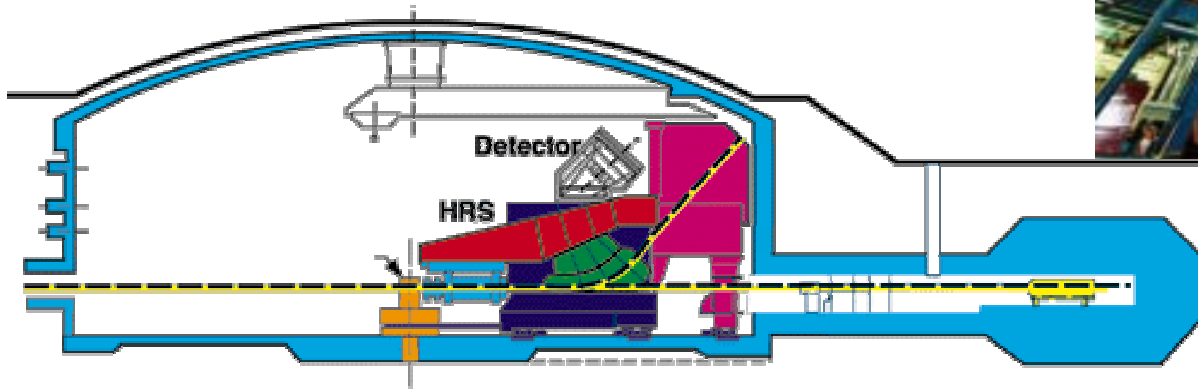
Stephen Wolbers, Jefferson Lab
Colloquium

16



Jefferson Lab

Hall A



Hall B

Hall C



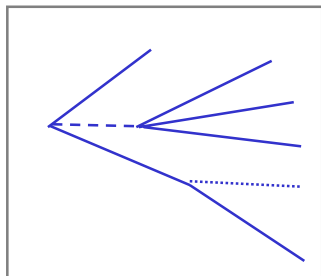
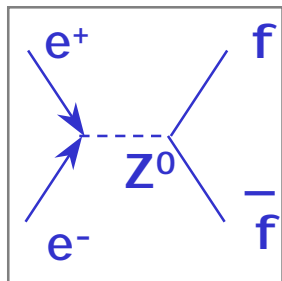
April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

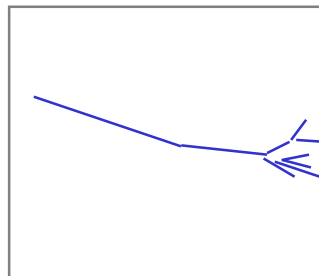
17



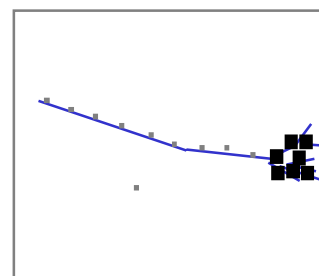
Physics to Raw Data (taken from Hans Hoffman, CERN)



Fragmentation,
Decay



Interaction with
detector material
Multiple scattering,
interactions



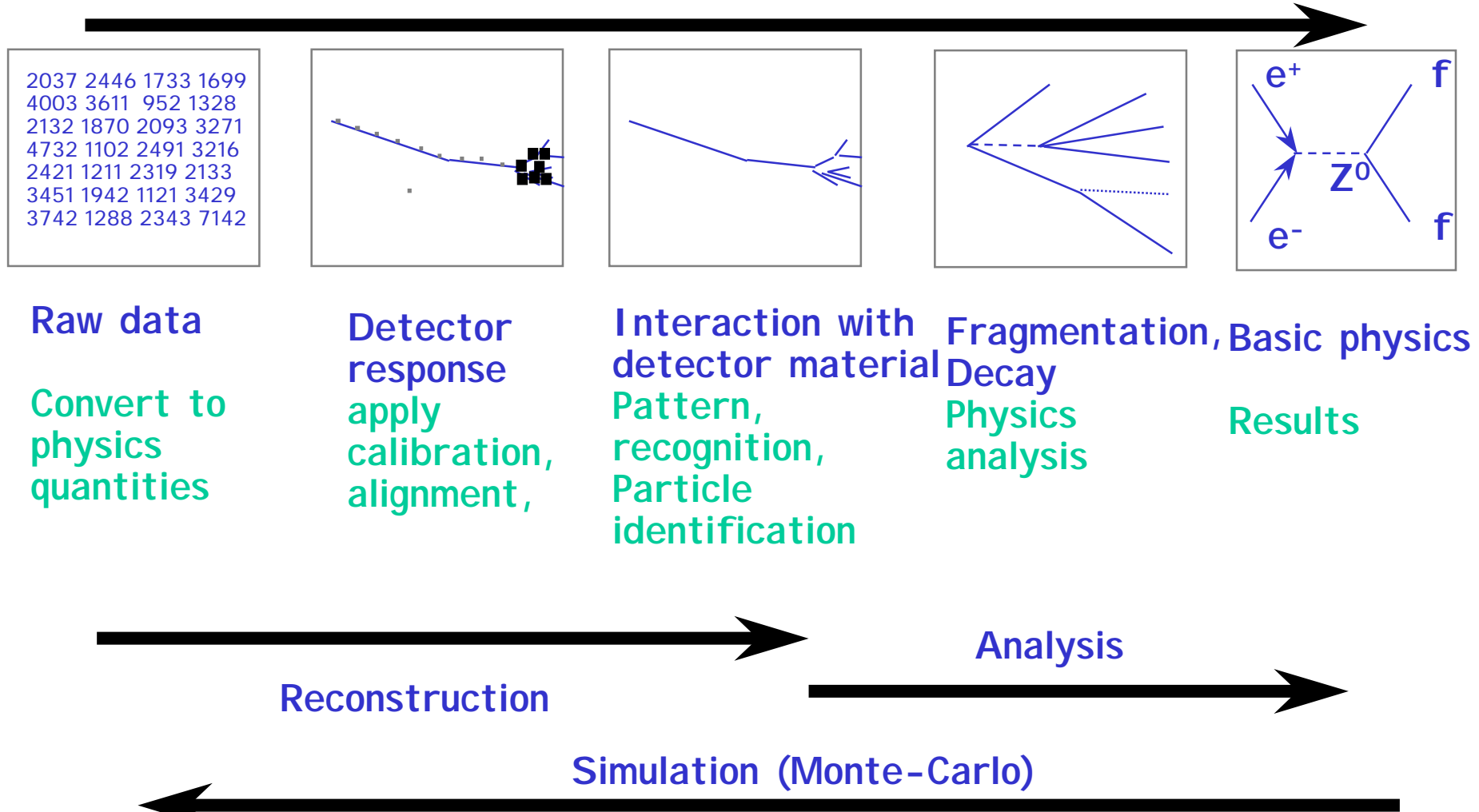
Detector
response
Noise, pile-up,
cross-talk,
inefficiency,
ambiguity,
resolution,
response
function,
alignment,
temperature

2037	2446	1733	1699
4003	3611	952	1328
2132	1870	2093	3271
4732	1102	2491	3216
2421	1211	2319	2133
3451	1942	1121	3429
3742	1288	2343	7142

Raw data
(Bytes)
Read-out
addresses,
ADC, TDC
values,
Bit patterns



From Raw Data to Physics





Computing Connection

Desired Improvement

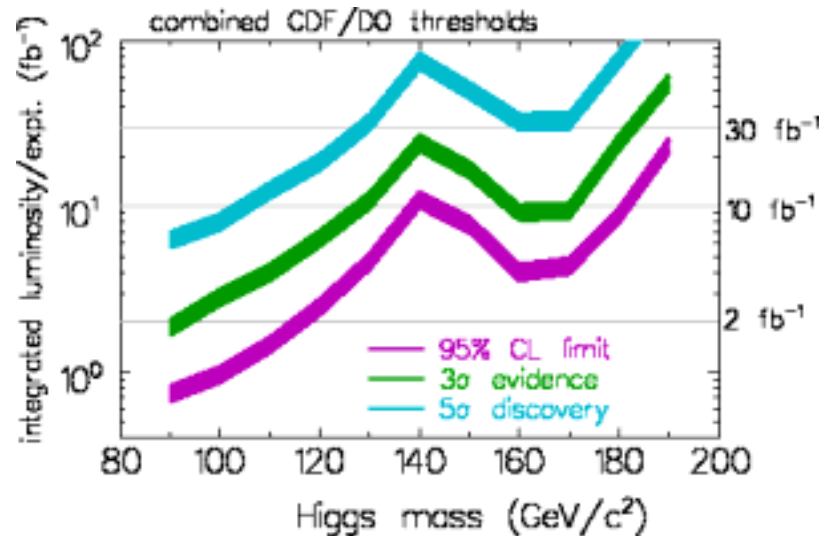
Computing Technique

Higher energy	—————→	Accelerator Design/simulation
More collisions	—————→	Acc. Design and controls
Better detectors	—————→	Triggers (networks, CPU), simulation
More events	—————→	Disk, tape, CPU, networks
Better analysis	—————→	Disk, tape, CPU, networks, algorithms
Simulation	—————→	CPU, algorithms, OO
Theory	—————→	CPU, algorithms, OO



Example – Higgs Search at CDF and D0 at the Tevatron

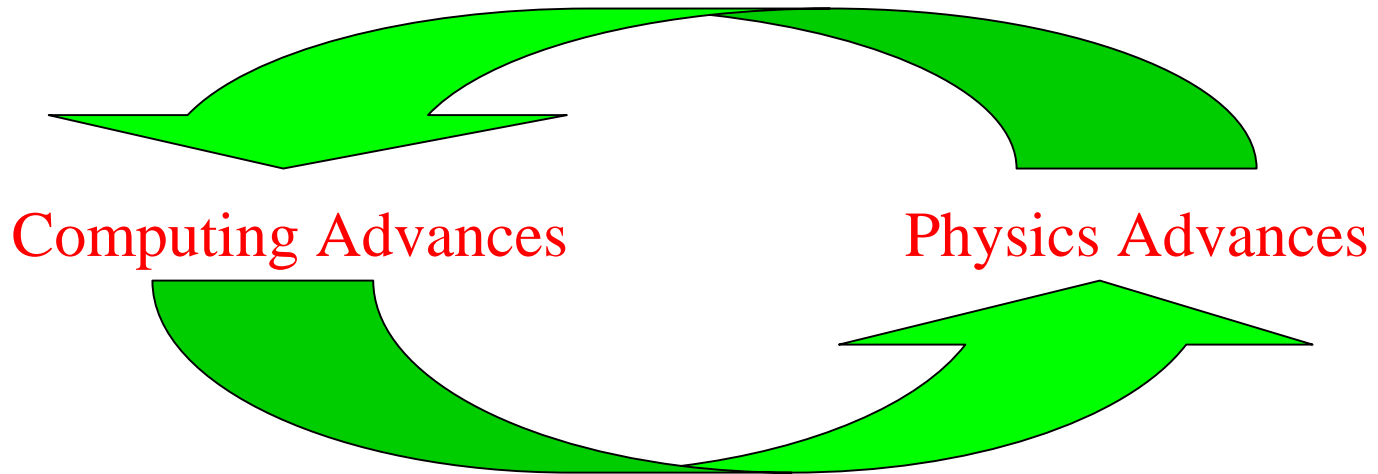
- Proton and Antiproton beam energy has been increased to 1 TeV (actually 980 GeV) and the luminosity of collisions has been (will be) dramatically increased
 - The pbar upgrades were modeled and tested on the Fermilab compute farms, as were the Recycler Ring design and Tevatron tuning.
 - Many of these calculations and tunings continue today.





Computing and Particle/Nuclear Physics Advances

- HENP has always required substantial computing resources
 - Computing advances have enabled “better physics”
 - Physics research demands further computing advances
 - Physics and computing have worked together over the years





Examples

- Data acquisition systems have advanced to allow increased data writing
 - Higher bandwidth DA systems
 - More sophisticated triggers, including software triggers
 - Even with the same beam and detector, an experiment can increase the “physics reach” by taking more data
- Increased computing power allows analysis of larger and larger datasets
 - This in turn allows for increased DA volume



Charm Physics

- Early experiments were able to study only a few hundred events.
- A series of experiments at Fermilab pushed to write a huge amount of data (for the time) in order to increase the statistics dramatically.
- This was possible because of the availability of large offline computing capability ("compute farms") to cope with the large data samples.



Data Volumes for HENP Experiments

(in units of 10^9 bytes)

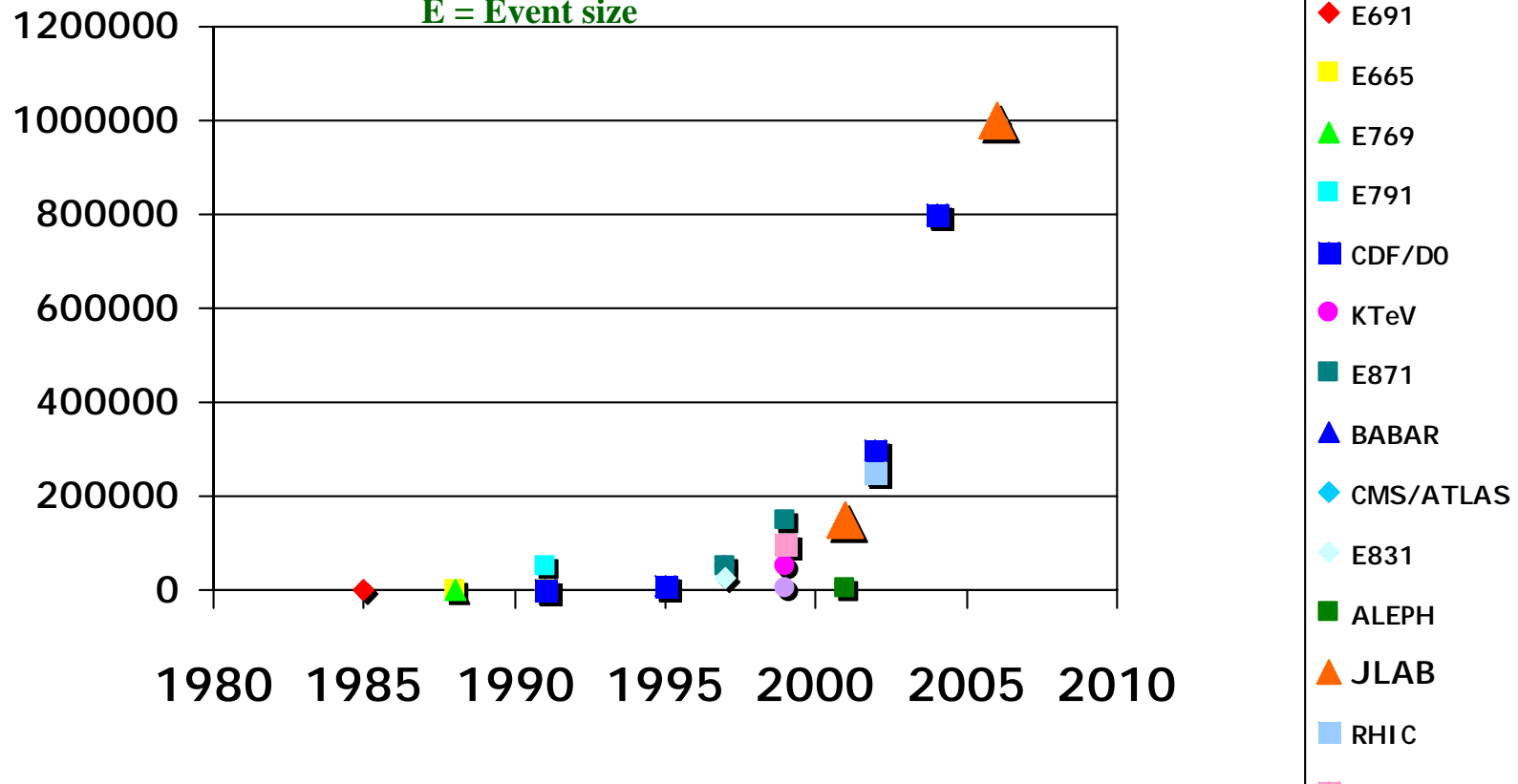
$$\text{Volume} = \sigma L \varepsilon E$$

σ = Cross Section

L = Luminosity

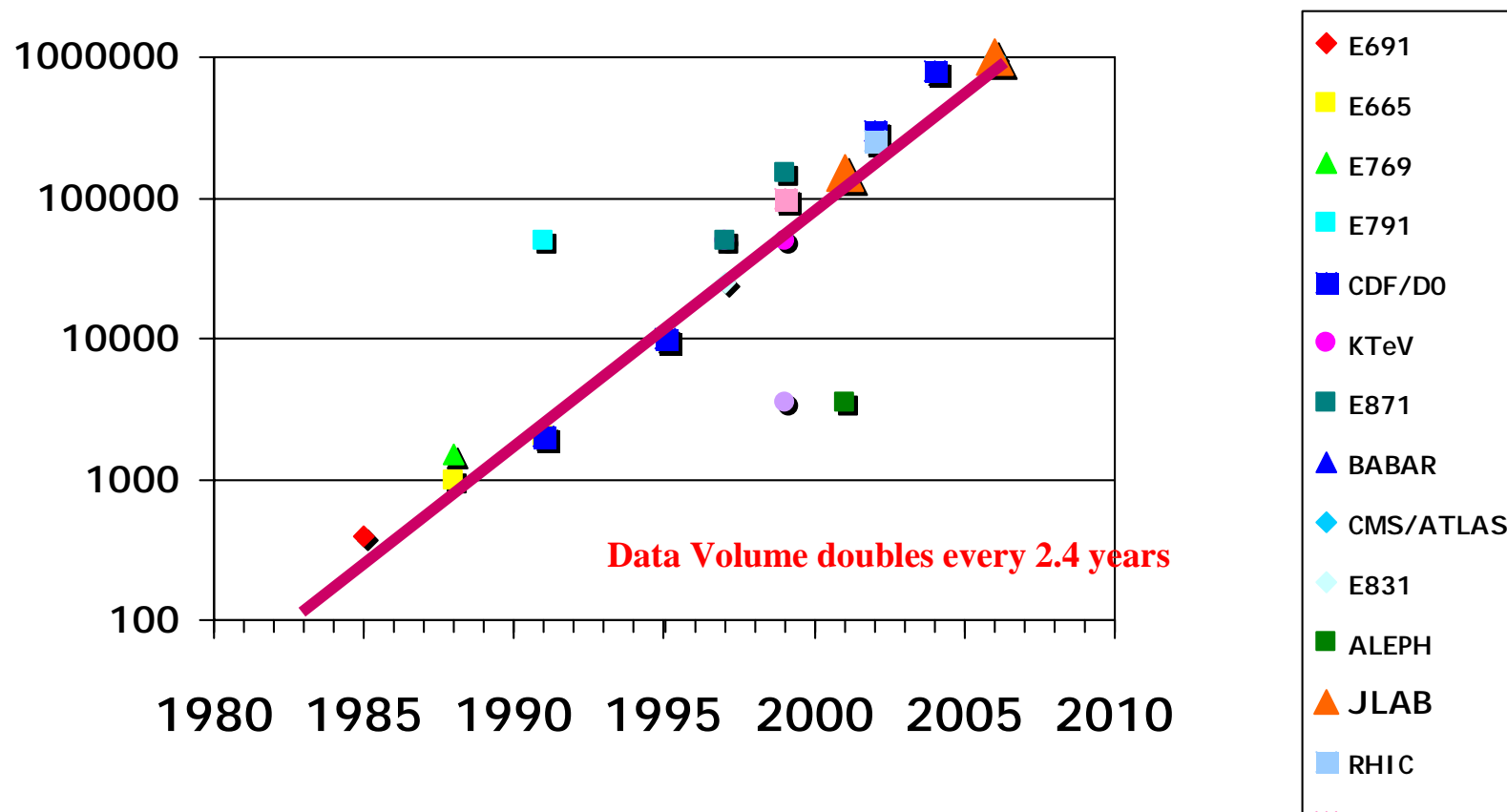
ε = Efficiency for collecting data

E = Event size





Data Volume per experiment per year (in units of 10^9 bytes)





Data Rates, Facilities and Techniques



Run 2a Computing

- Data Rates and Volume
- Software Development
- Computing Model
 - Event Reconstruction
 - Analysis
 - Mass Storage



Run 2a Computing

- **Scope of Computing Problem**
 - Data volume of ~1 Pbyte (over 2 years, CDF+D0)
(Pbyte = 10^{15} bytes = 1,000,000 Gbytes)
 - Typical hard disk is 20 Gbytes, DVD is 5 Gbytes.
 - Rates out of detector to storage up to 20 Mbyte/sec (each detector).
 - CPU for event reconstruction of about 5 sec/event on a PIII /500 MHz PC (Each event is 250 Kbyte).
 - CPU for data analysis is supplied on large SMP systems and Linux desktops.
 - Large disk arrays (>50 Tbyte).
 - 100's of physicists spread across the world need to analyze the data.



Software Development



Software Development

- **Event Reconstruction Software**
 - Written by physicists.
 - Translates detector output (ADC counts, TDC, hit maps) into energy measurements, particle positions and directions and momentum.
 - Written in FORTRAN in previous runs.
 - Written in C++ in Run 2.
 - Hundreds of packages or modules, millions of lines of code, many 10's of authors.



C++ Experience

- Big change from procedural to object-oriented language.
- Some resistance.
- Large training requirements.
- Need for C++ experts to support the physicists on design and coding.
 - Two individuals were hired by Fermilab to provide that support.
- The code runs, is probably as fast or faster than Fortran code, and in general the exercise has been successful.
- Most (not all) new experiments choose C++ for offline event reconstruction.



Was the transition to C++ beneficial?

- I'm not an expert and haven't worked with the code directly.
- The answer probably won't be known for some time:
 - Will code be more easily maintainable?
 - Will the code be more robust?
 - Will the code be as fast or at least not too slow?
 - Will we be aligned better with industry and other code developers?



Other software for Run 2

- Mixture of commercial, lab-developed and open source.
- Each product is chosen based on its ability to solve a problem and on its cost (both to write and to support).
- Long list of products, some examples:
 - Linux, gcc, emacs, MySQL
 - KAI C++ compiler, LSF (Batch system), Purify
 - FBS, Enstore, SAM, ftt, ZOOM
 - GEANT3/4, ROOT



Computing Model



Computing Model for Run 2a

- CDF and D0 have similar but not identical computing models.
 - In both cases data is logged to tape stored in large robotic libraries.
 - Event reconstruction is performed on large Linux PC farms.
 - Analysis is performed on medium to large multi-processor computers
 - Final analysis, paper preparation, etc. is performed on Linux desktops or Windows desktops.

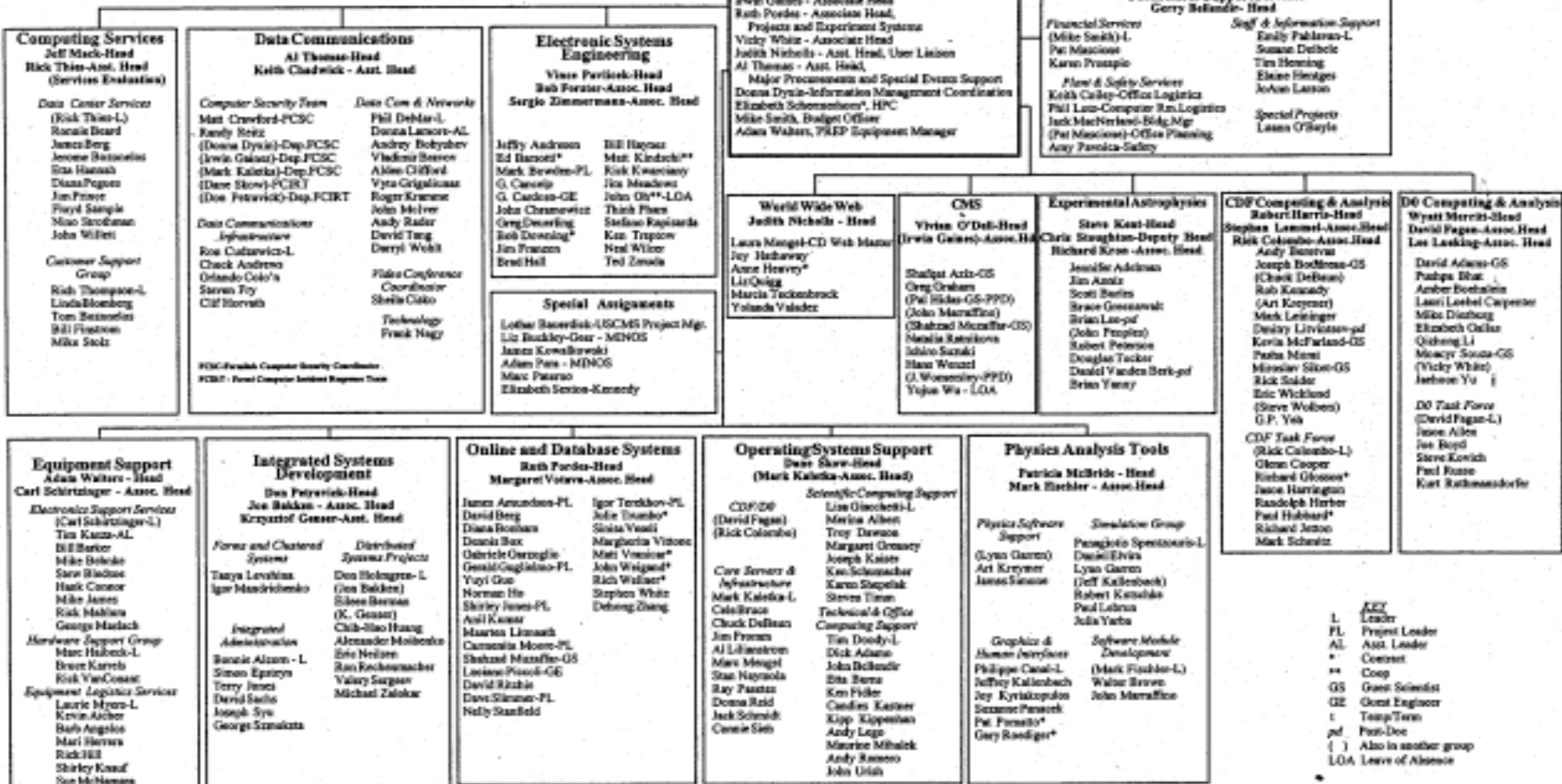


Support for Run 2 Computing

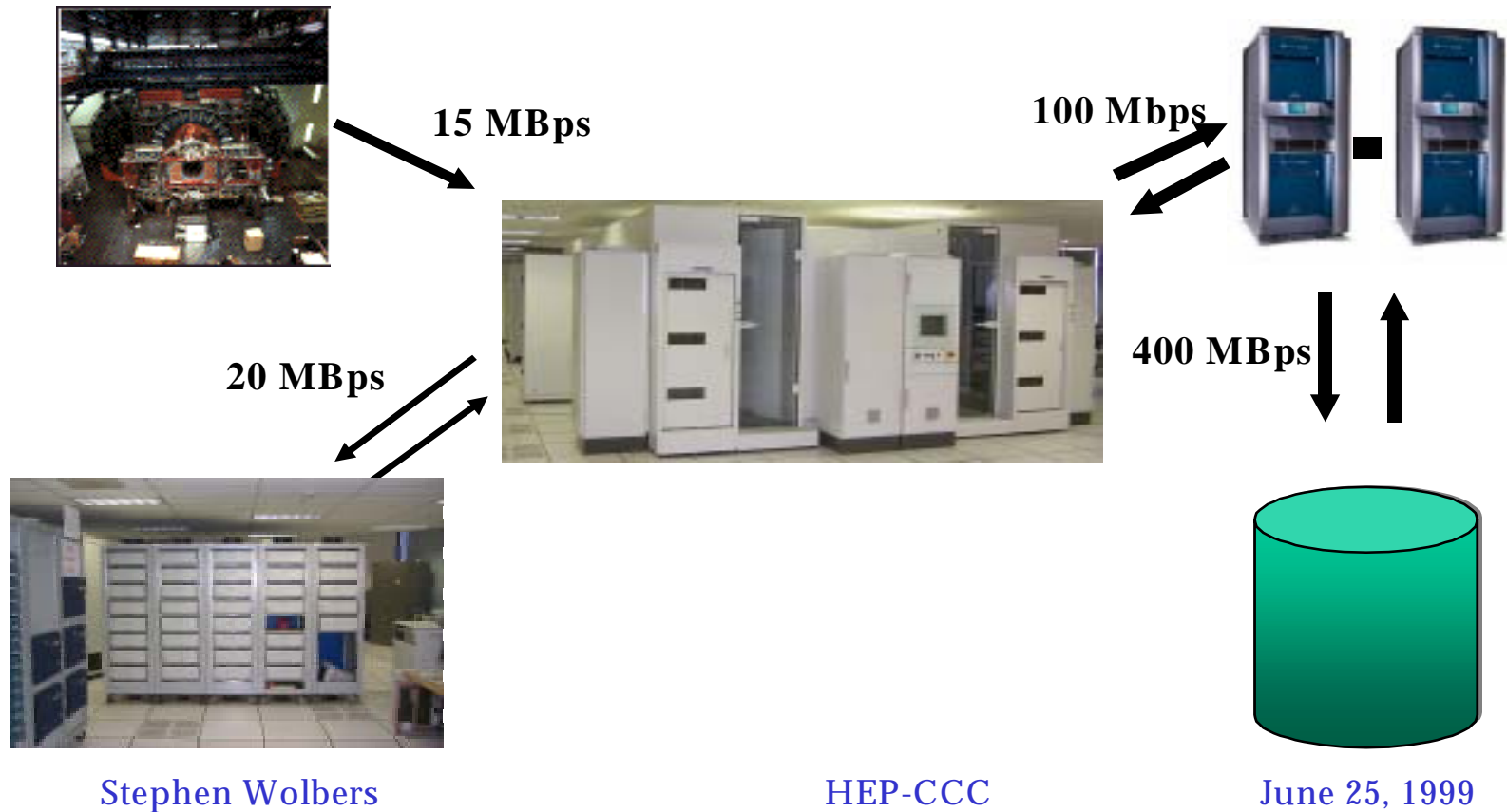
- It takes many people to build and support Run 2 computing systems
 - Design and build: 20-30 people per experiment
 - This includes code development and systems integration. Other support like networks, system administration is not included here.
 - Maintain: About 35 people per experiment
 - This includes product support, networks, system administration, etc.
 - Some of this support is provided by scientists from the two experiments.



Computing Division Organization Chart APRIL 9, 2001



RunII Data Flows



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium



Reconstruction Systems



CDF/D0 Offline Production Farms for event reconstruction

- The CDF/D0 farms must have sufficient capacity for Run 2 Raw Data Reconstruction.
- The farms also must provide capacity for any reprocessing needs.
- Farms must be easy to configure and run.
- The bookkeeping must be clear and easy to use
- Error handling must be excellent.

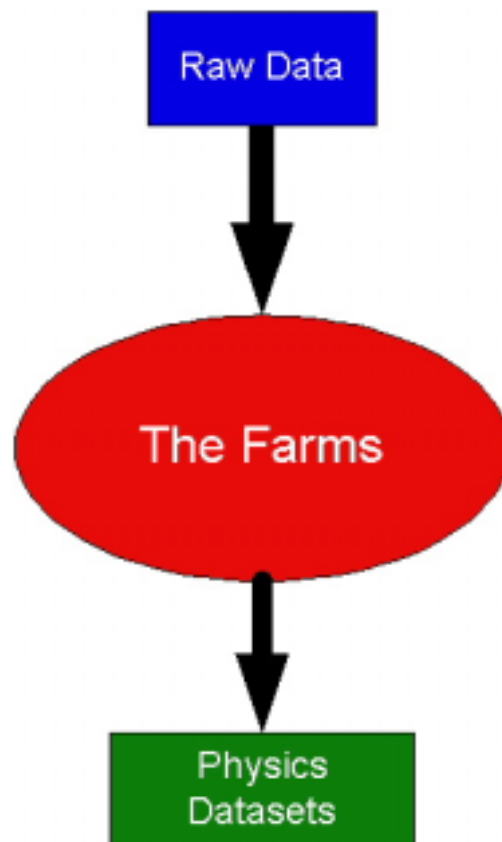


Requirements for CDF (CPU)

- CPU goal is <5 seconds/event on PIII/500
- Assuming 70% efficiency this translates to
 - 200 PIII/500 equivalents (each is about 20 SpecInt95)
 - 4200 SpecInt95
- Adding in reprocessing, simulation, responding to peak rates
 - 300-400 PIII/500 equivalents (150-200 duals)
 - 6300-8400 SpecInt95

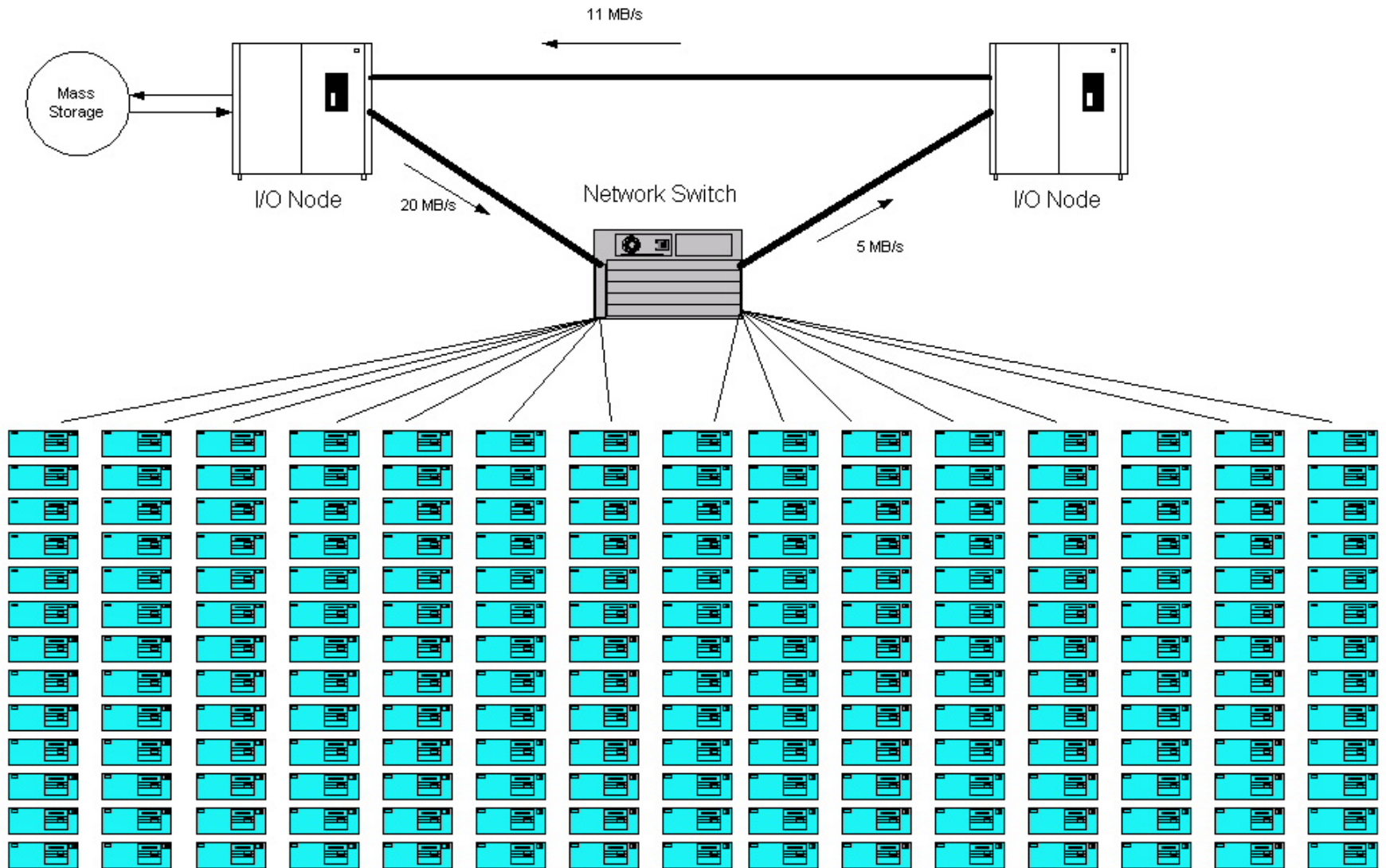


Simple Model





Run II CDF PC Farm





Design/Model

• Hardware

- Choose the most cost-effective CPU's for the compute-intensive computing.
- This is currently the dual-Pentium architecture
- Network is fast and gigabit ethernet, with all machines being connected to a single or at most two large switches.
- A large I/O system to handle the buffering of data to/from mass storage and to provide a place to split the data into physics datasets.



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

46



April 25, 2001



Stephen Wolbers, Jefferson Lab
Colloquium

47

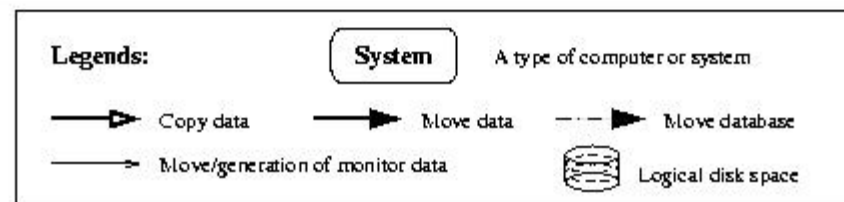
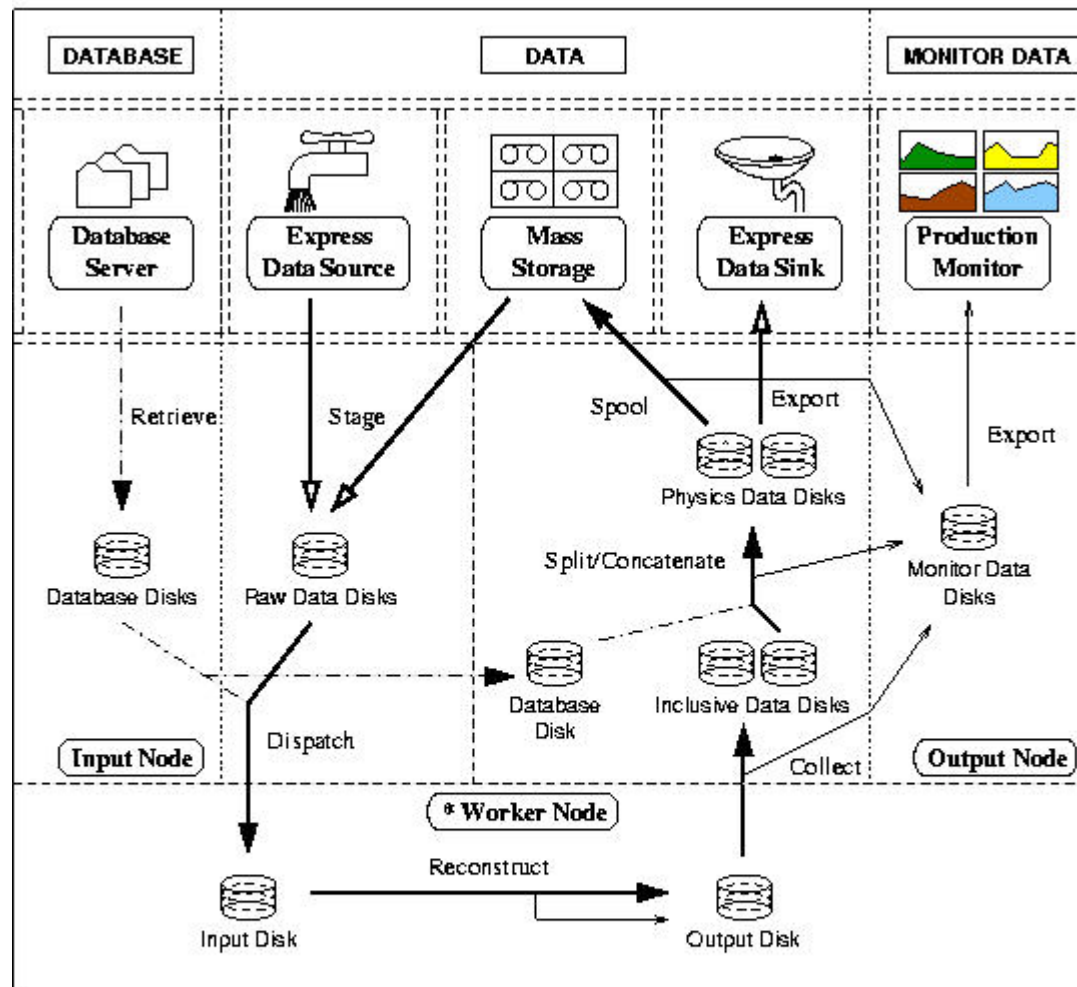


Software Model

- **Software consists of independent modules**
 - Well defined interfaces
 - Common bookkeeping
 - Standardized error handling
- **Choices**
 - Python
 - MySQL database (internal database)
 - FBSNG (Farms Batch System)
 - FI PC (Farms Interprocessor Communication)
 - CDF Data Handling Software



Conceptual Model of Run 2 Production System



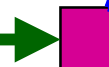
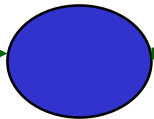
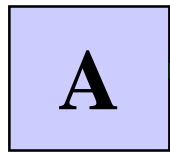


Physics Analysis Requirements and Impact

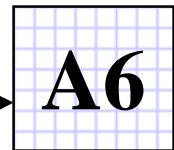
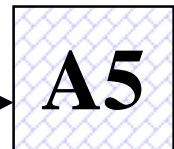
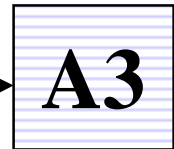
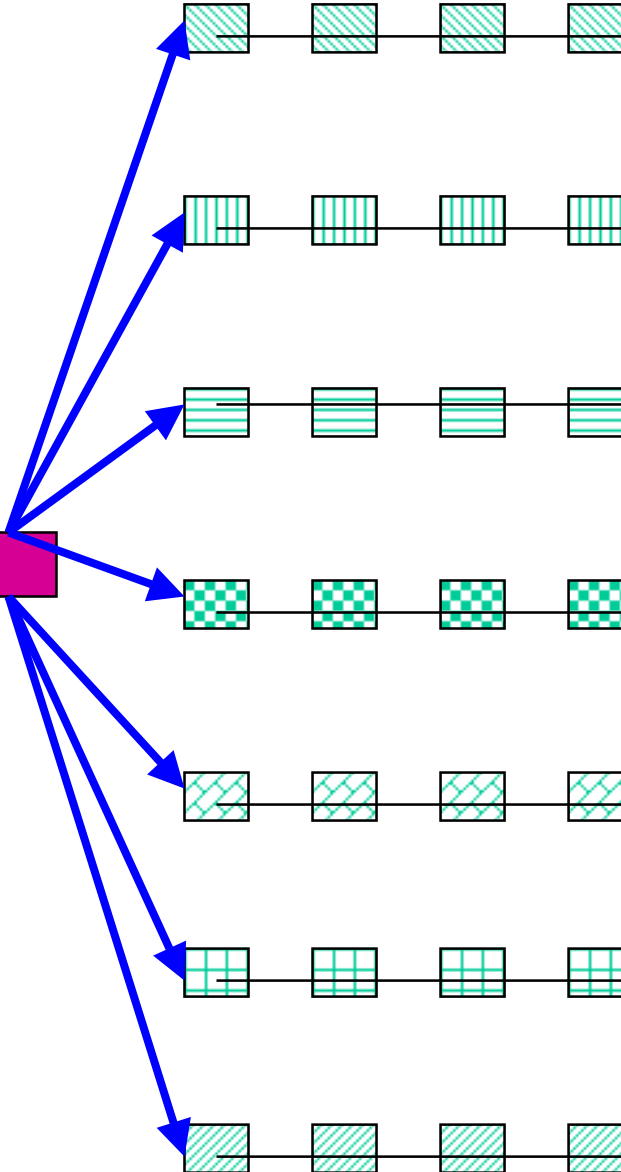
- Raw Data Files come in ~8 flavors, or streams
 - 1 Gbyte input files
- Reconstruction produces inclusive summary files
 - 250 Mbyte output files
- Output Files must be split into ~8 physics datasets per input stream
 - Target 1 Gbyte files
 - About 20% overlap
- Leads to a complicated splitting/concatenation problem, as input and output streams range from tiny (<few percent) to quite large (10's of percent)



Input Stream (x8)



Farms



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium



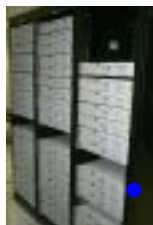
Status of CDF/D0 Farms

- 88+97 PC's are in place.
 - 48+47 PIII /500 duals
 - 40+53 PIII /750 or 800 duals
- 60 more PC's are on order (PIII /1 GHz duals)
- I/O nodes are ready.
- Integration and testing of the system is complete.
 - 20 Mbyte/sec can be achieved.
- The CDF system is being used to process and reprocess data from the commissioning run (about 1.3 Tbytes taken in October, 2000) and both systems are used to generate and reconstruct simulated data.
- Both are ready for raw data reconstruction.



Analysis Computing

Analysis Computing – Run 2a



- CDF and D0 have both acquired large Silicon Graphics O2000 multi-processor systems for large analysis tasks.



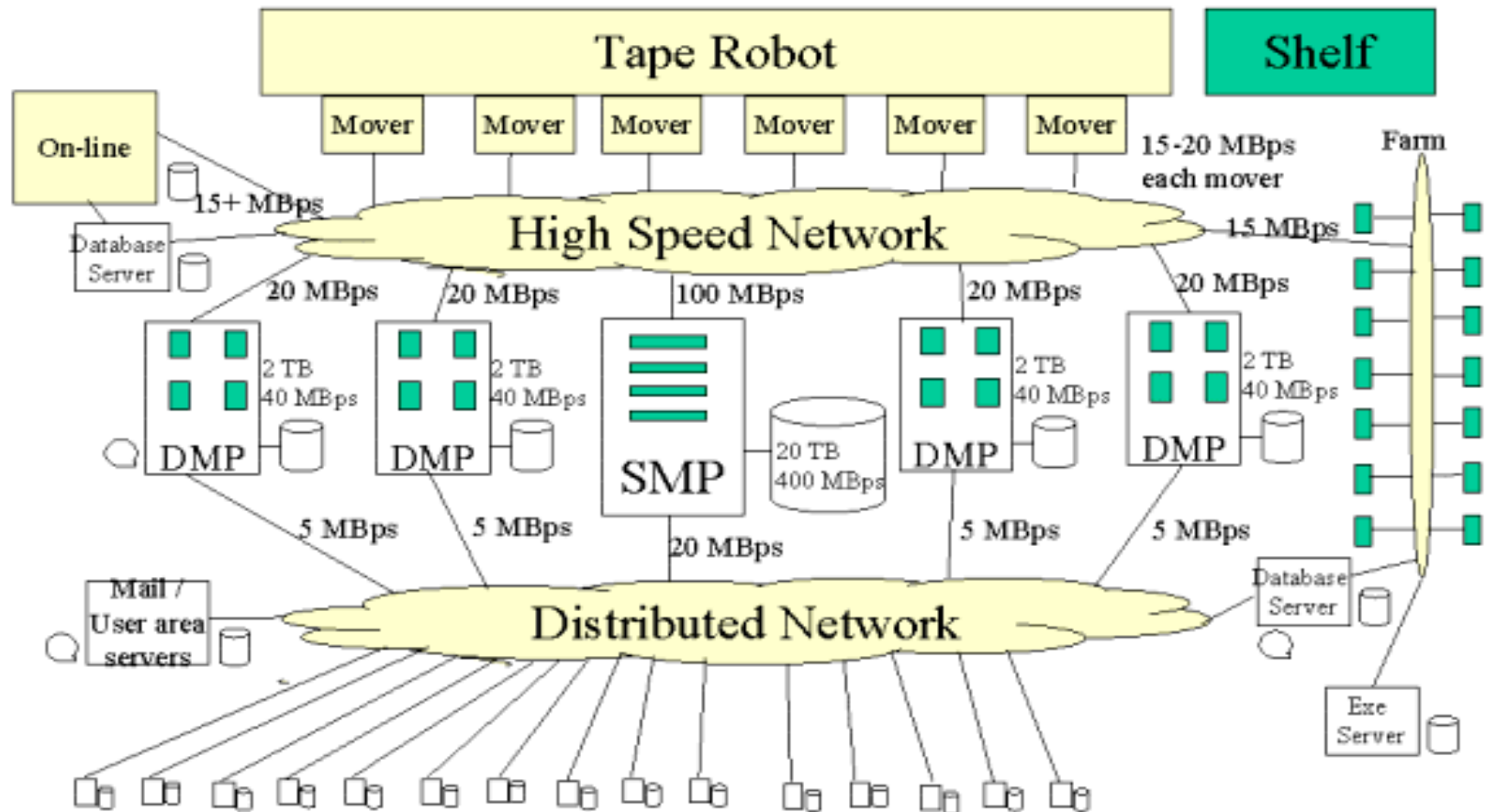


Analysis Computing

- Each system has:
 - Access to data on tape.
 - D0 access is over the network
 - CDF access is via SCSI connected peripherals
 - Access to disk storage.
 - About 30 Tbytes attached to central systems
 - This will increase, especially as disk prices continue to fall
 - LSF Batch software is used to schedule jobs and manage resources on these systems.



Proposed D0 Analysis Computing Configuration



400 desktop : 50GB, 0.1MBps (avg), 10 MBps (burst) ea

SMP = Symmetric Multi-processor

○ Tape Backup

DMP = Distributed multi-processor

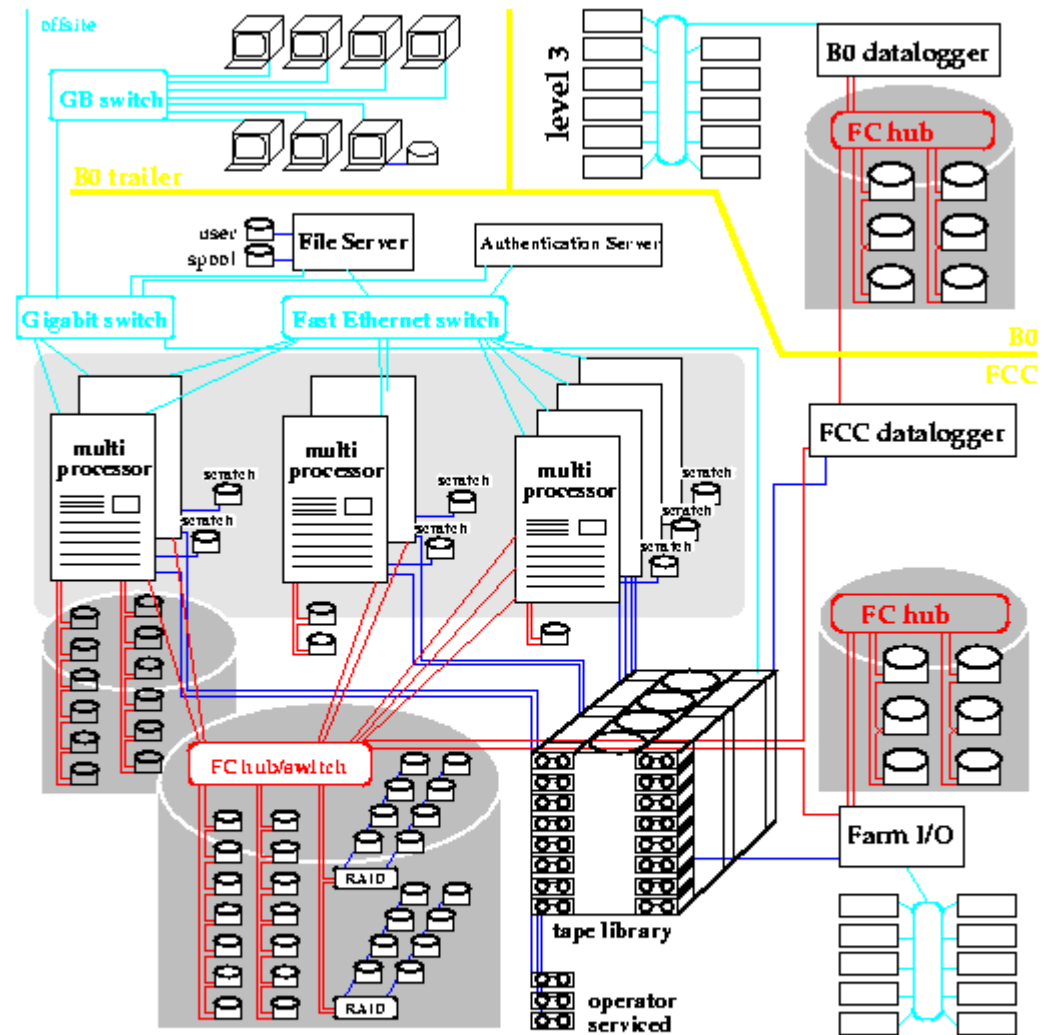
9/21/98



Data Access Model: CDF

- Ingredients:**

- Gigabit Ethernet
- Raw data are stored in tape robot located in FCC
- Multi-CPU analysis machine
- High tape access bandwidth
- Fiber Channel connected disks





Analysis Computing

- The large SGI is a conservative (and expensive) solution to analysis computing needs.
- Both collaborations are exploring the use of PC's + EIDE disk + 100 Mbit or 1 Gbit network connection for analysis.
- These projects may lead the way to more cost-effective solutions for the analysis of the large amount of data that will be taken in Run 2a and Run 2b.



PC analysis computing -- examples



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

59



Mass Storage



Storing Petabytes of Data in mass storage

- Storing (safely) petabytes of data is not easy or cheap.
 - Need large robots (for storage and tape mounting).
 - Need many tapedrives to get the necessary I/O rates.
 - Tapedrives and tapes are an important part of the solution, and has caused some difficulty for Run 2.
 - Need bandwidth to the final application (network or SCSI).
 - Need system to keep track of what is going on and schedule and prioritize requests.



Robots and tapes



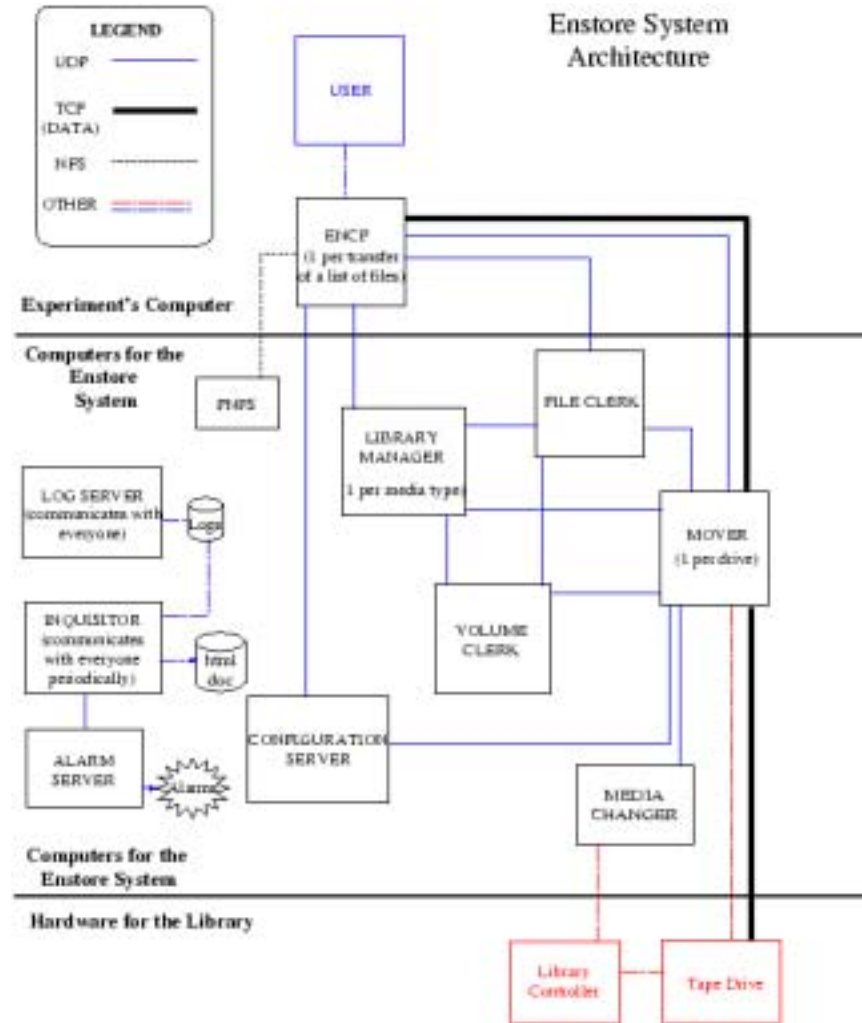
April 25, 2001

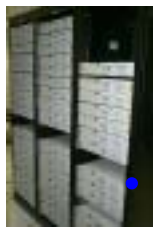
Stephen Wolbers, Jefferson Lab
Colloquium

62



Enstore Software System



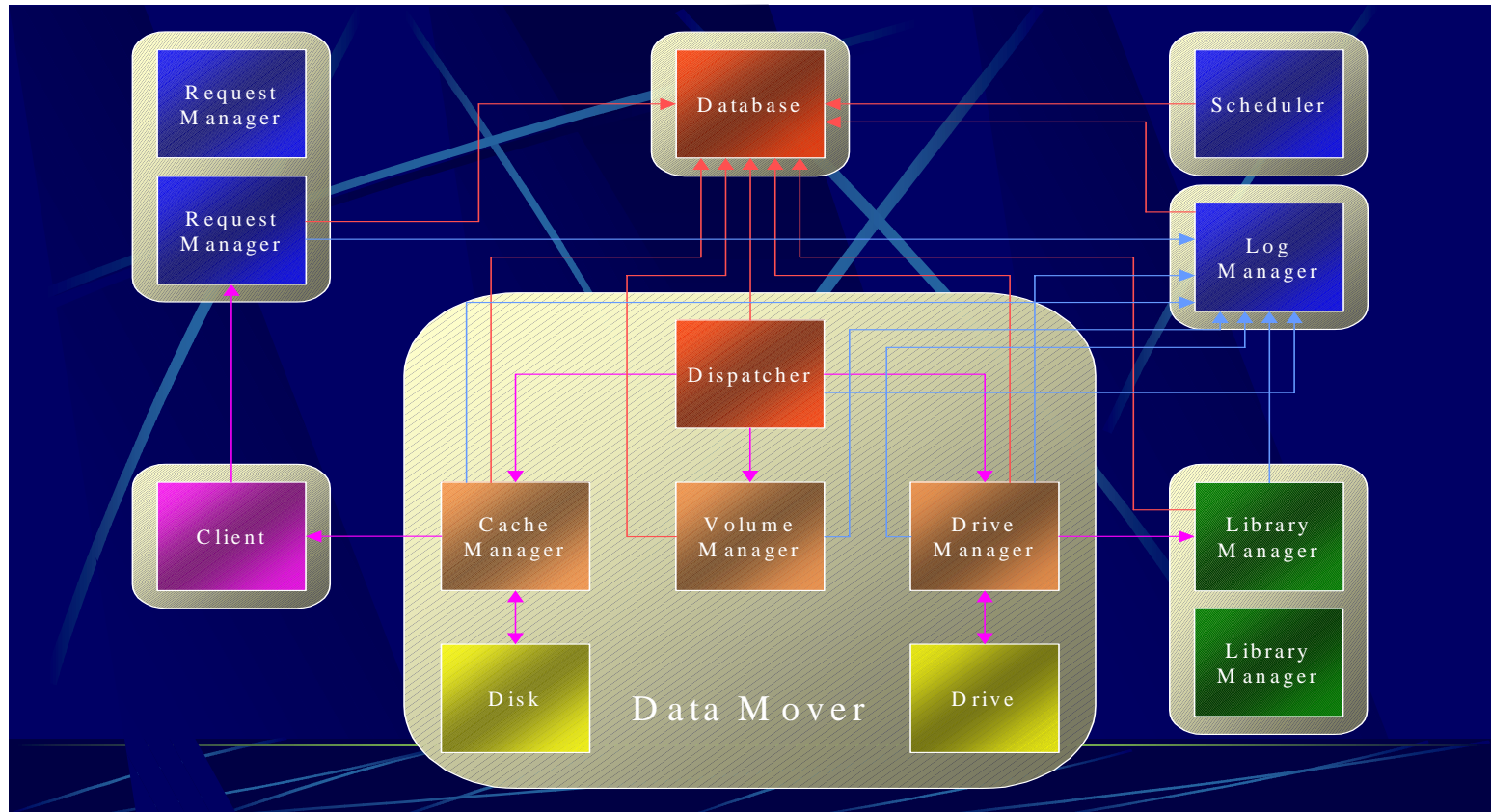


Tapedrives and tapes

- Tapedrives are not always reliable, especially when one is pushing for higher performance at lower cost.
- Run 2 choice is Exabyte Mammoth 2.
 - 60 Gbytes/tape.
 - 12 Mbyte/sec read/write speed.
 - About \$1 per Gbyte for tape. (A lot of money.)
 - \$5000 per tapedrive.
- AIT2 from SONY is the backup solution.
- The robotics which exist can handle most any tapedrive technology.
- Given the Run 2 timescale, upgrades to newer technology will occur.
- Finally, Fermilab is starting to look at PC diskfarms to replace tape completely.

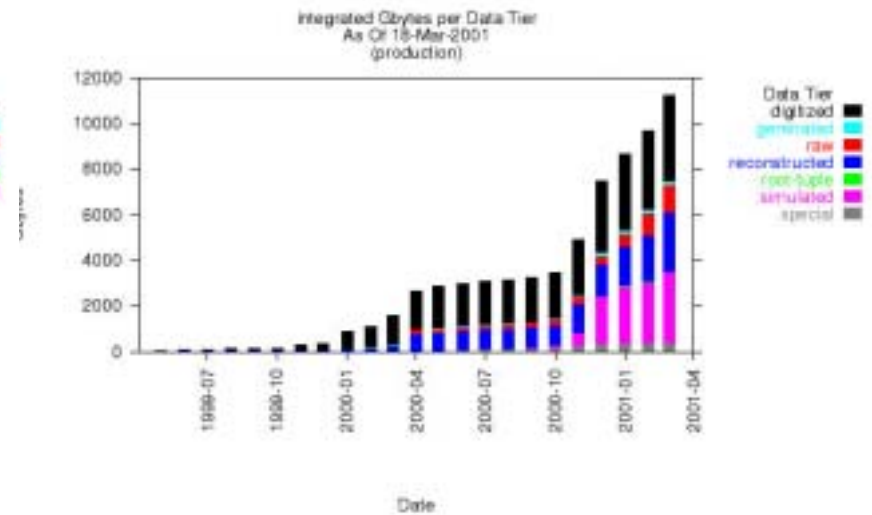
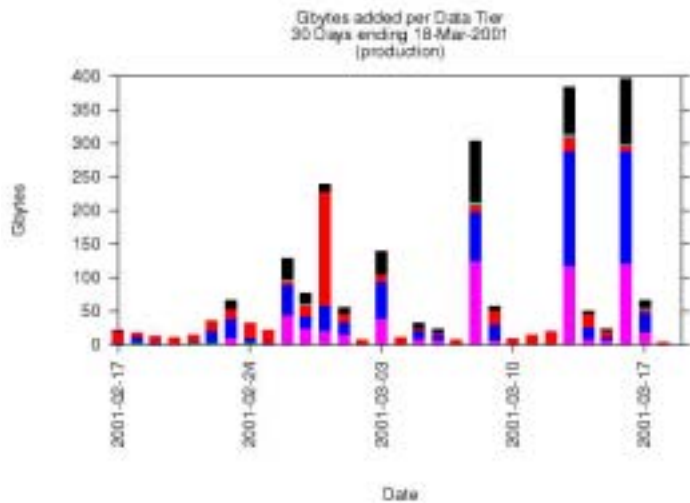


Software system, Jasmine, (from Jefferson Lab)





Data Flow into/out of mass storage (D0)





RHI C and JLAB Computing

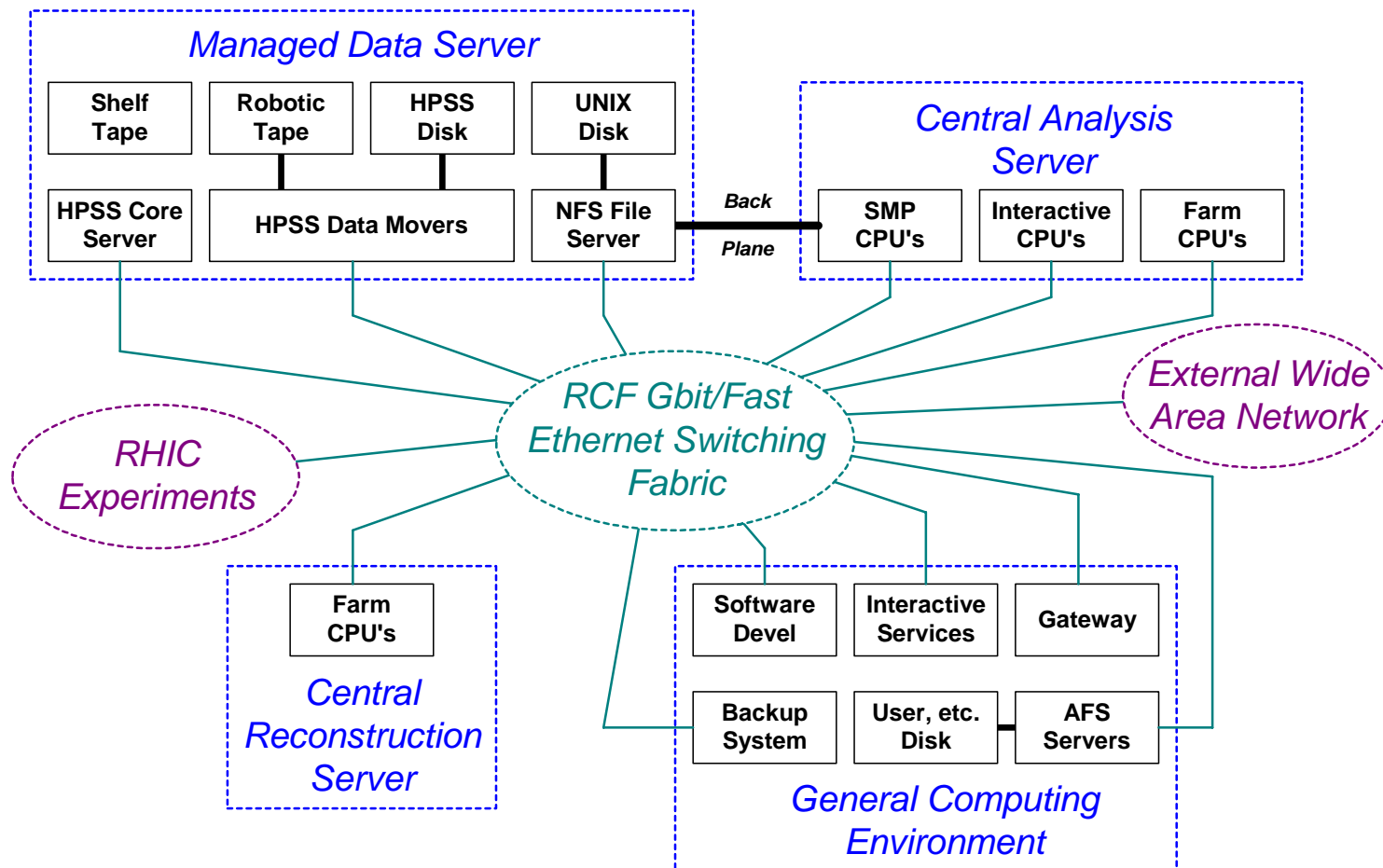


RHI C Computing

- Computing problem is very similar and is in fact somewhat larger than the Run 2a computing problem.
- For Nominal Year Operations, *2001/2002*
 - Aggregate Raw Data Recording at 60 MBytes/sec
 - Annual Data Storage: 1 PByte
 - Online Storage: 40 TBytes
 - Online Data Access at 1 GByte/sec
 - Installed Compute Capacity: 20,000 SPECint95



RHI C Computing Facility





April 25, 2001

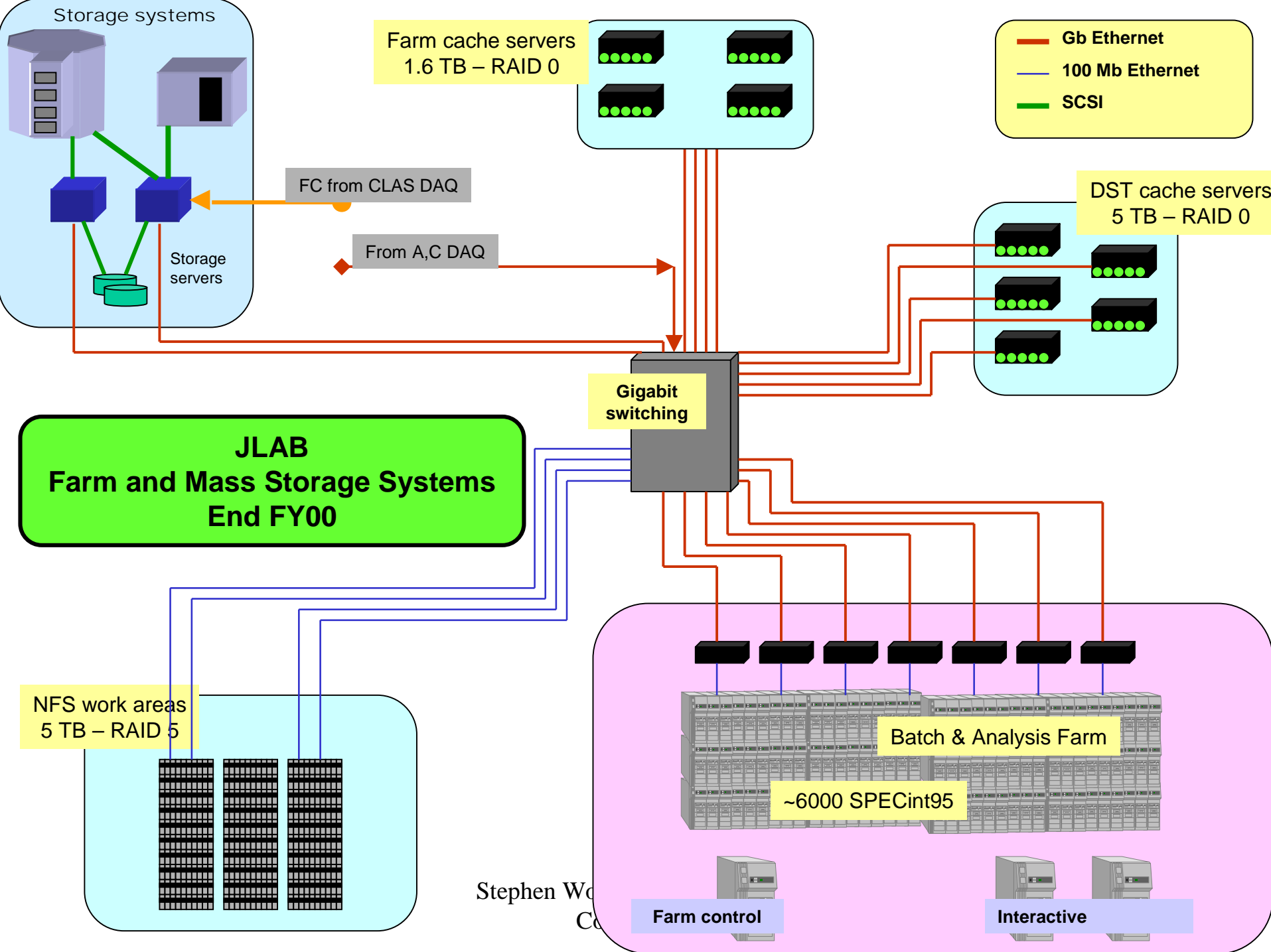
Stephen Wolbers, Jefferson Lab
Colloquium

70



Jefferson Lab Computing

- Jefferson Lab Computing Problem is also quite substantial
 - Data rates are comparable to CDF/D0/RHIC.
 - Need large reconstruction, storage, and analysis systems.
 - Future increases (for new beams and facilities) will have to be handled.





Lattice QCD



Lattice QCD

- Lattice QCD is a powerful approach to study QCD and to calculate fundamental quantities in the theory.
- Lattice QCD calculations require extremely large computing power in a tightly-coupled computing architecture (because of the demand for fast, low-latency communications).
- The computers used for this are almost always special-purpose machines, designed for this class of calculations
- Examples include, APE, Columbia machine, ACPMAPS, etc.



Commodity Lattice Gauge Machine

- Idea: Take advantage of commodity hardware and software to build a large lattice QCD machine.
- Goal: 10 Teraflop peak performance as cheaply as possible.
- R&D at Fermilab (similar work at Jefferson lab).
 - Small Machine (80 dual PC's) has been purchased and is being integrated at Fermilab.
 - Much larger machine (on order 1000 PC's) will be built assuming that funding is available, no serious problems are found in scaling, etc.
 - Workshop was held March 26-28 at Fermilab to discuss the current ideas and progress.



Old and New Lattice Gauge Computing at Fermilab



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium

76



Accelerator Physics

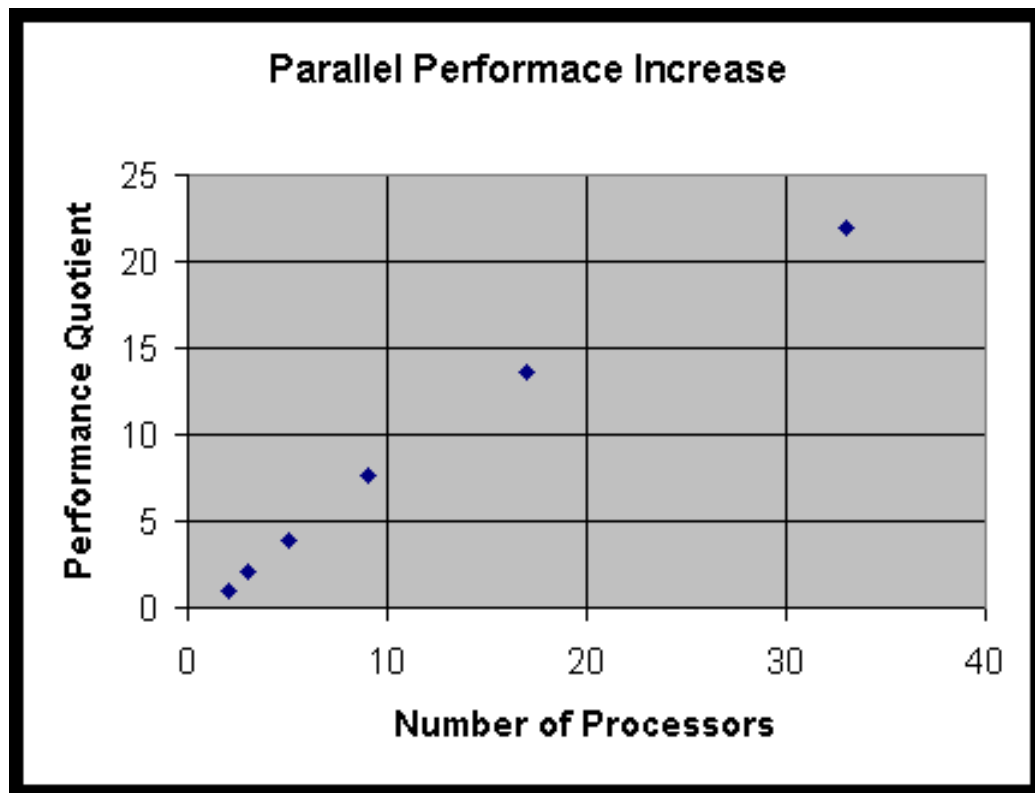
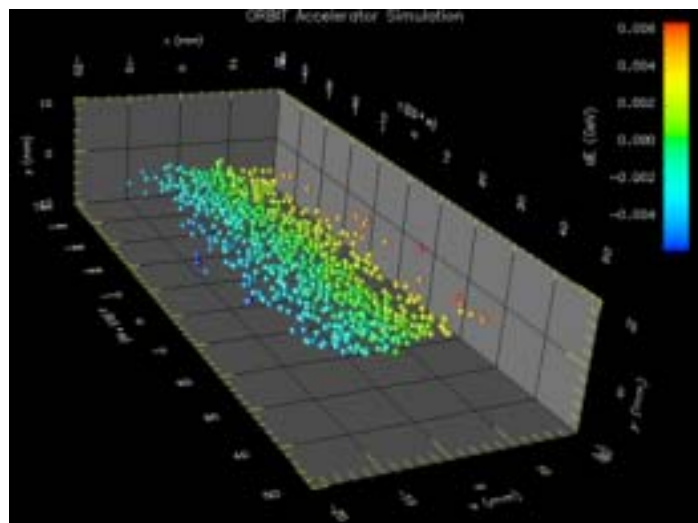


Accelerator Design and Tuning

- Computing is essential for accelerator design and tuning:
 - Optimizing design of future accelerators:
 - Electron-positron linear collider
 - Muon collider/neutrino factory
 - Very Large Hadron collider
 - Getting the most out of current machines:
 - Higher luminosity
 - With lower backgrounds/halo, etc
 - Parallel computing can potentially decrease the time taken to properly simulate each configuration and allow for faster and more complete design.



Accelerator Design and Computing





Future



Run 2b at Fermilab

- Run 2b will start in 2004 and will increase the integrated luminosity to CDF and D0 by a factor of approximately 8 (or more if possible).
- It is likely that the computing required will increase by the same factor, in order to pursue the physics topics of interest:
 - B physics
 - Electroweak
 - Top
 - Higgs
 - Supersymmetry
 - QCD
 - Etc.



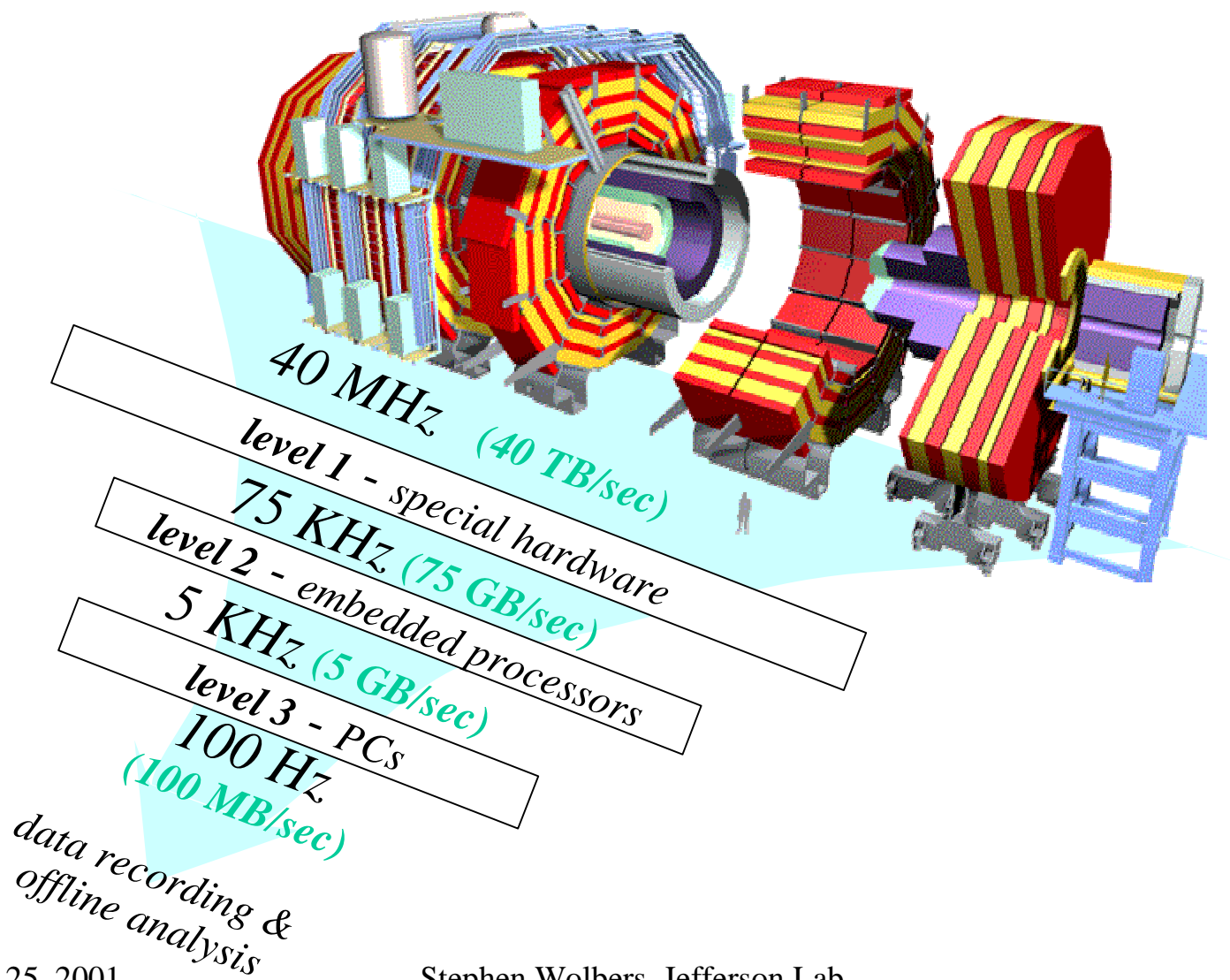
Run 2b Computing

- **Current estimates for Run 2b computing:**
 - 8x CPU, disk, tape storage.
 - Expected cost is same as Run 2a because of increased price/performance of CPU, disk, tape.
 - Plans for R&D testing, upgrades/acquisitions will start next year.
- **Data-taking rate:**
 - May be as large as 80 Mbyte/s.
 - About 1 Petabyte/year to storage.



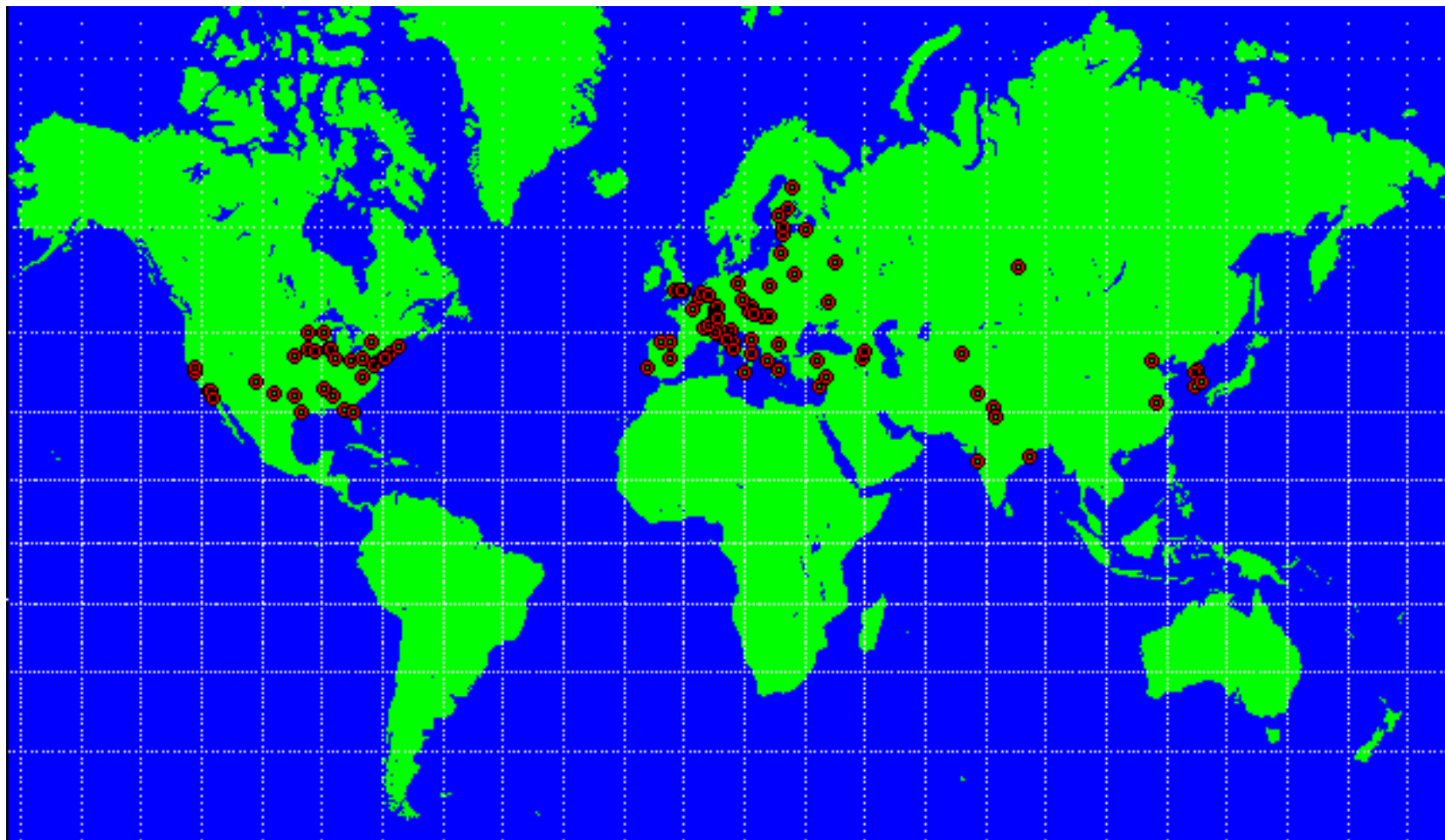
LHC Computing

- LHC (Large Hadron Collider) will begin taking data in 2006–2007 at CERN in Switzerland.
- Data rates per experiment of 100 Mbytes/sec.
- 1 Pbyte/year of storage for raw data per experiment.
- World-wide collaborations and analysis.
 - Desirable to share computing and analysis throughout the world.
 - GRID computing may provide the tools.



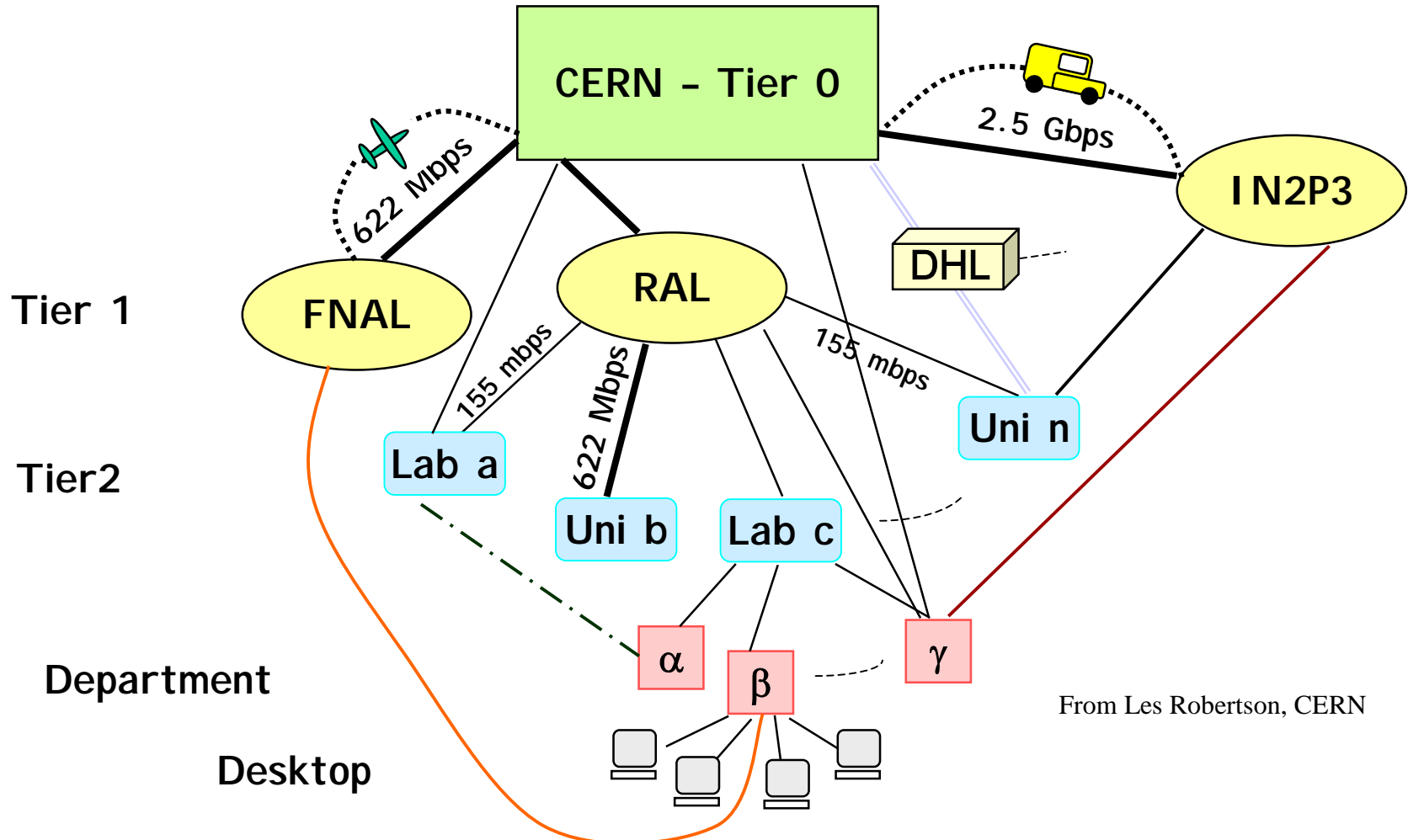


World Wide Collaboration
 \Rightarrow *distributed computing & storage capacity*



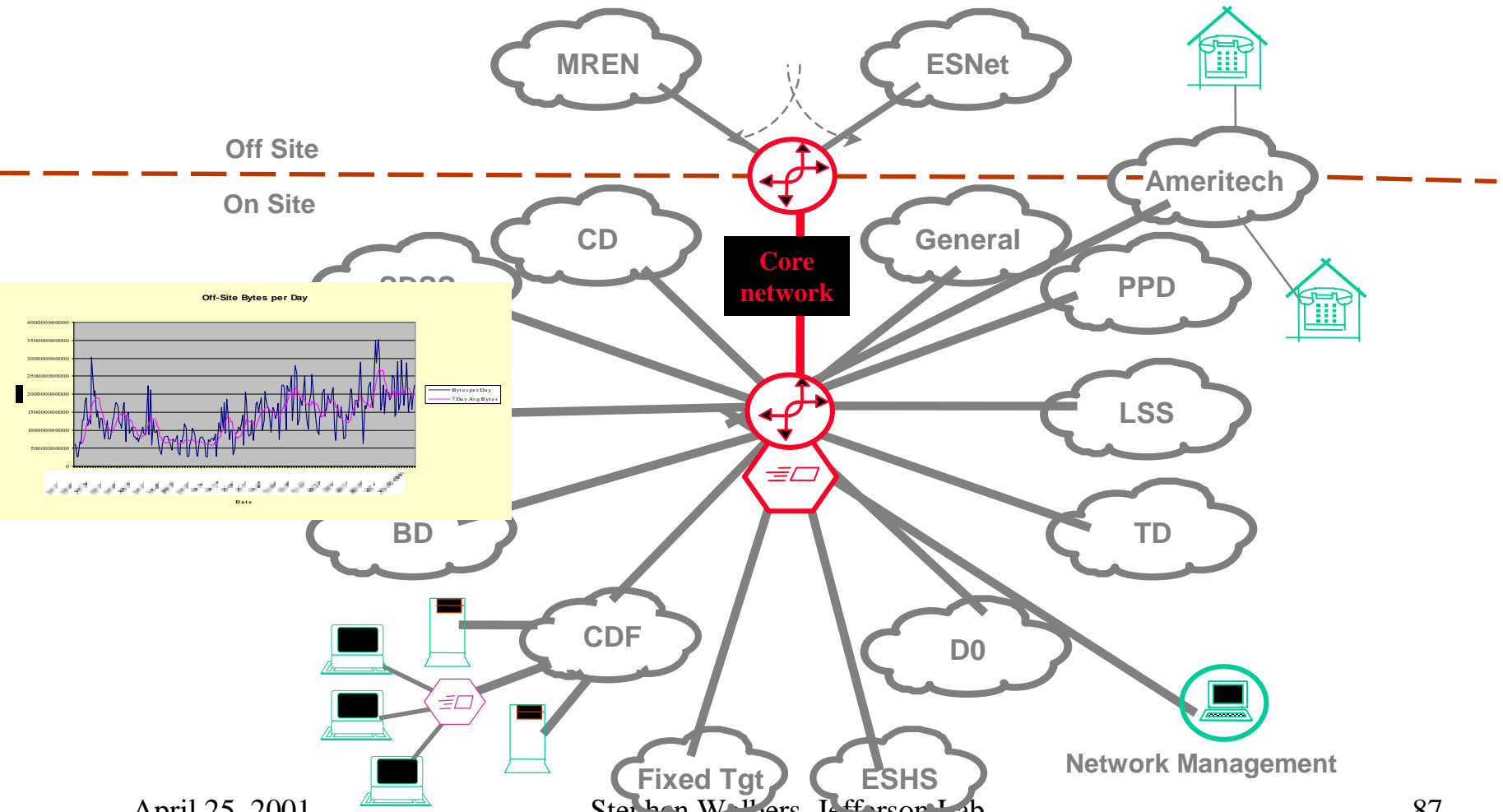


CMS/ATLAS and GRID Computing



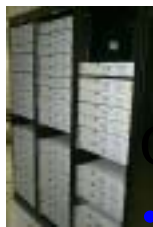


Fermilab Networking and connection to Internet



April 25, 2001

Stephen Wolbers, Jefferson Lab
Colloquium



Are Grids a solution?

Computational Grids

Les Robertson, CERN

- Change of orientation of Meta-computing activity
 - From inter-connected super-computers
... .. towards a more general concept of a
computational power Grid (The Grid – Ian Foster,
Carl Kesselman^{**})
- Has found resonance with the press, funding
agencies

But what is a Grid?

*"Dependable, consistent, pervasive access to
resources^{**}"*

So, in some way Grid technology makes it easy to use
diverse, geographically distributed, locally managed
and controlled computing facilities – as if they
formed a **coherent local cluster**

^{**} Ian Foster and Carl Kesselman, editors, "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann, 1999



What does the Grid do for you?

Les Robertson

- You submit your work
- And the Grid
 - Finds convenient places for it to be run
 - Organises efficient access to your data
 - Caching, migration, replication
 - Deals with authentication to the different sites that you will be using
 - Interfaces to local site resource allocation mechanisms, policies
 - Runs your jobs
 - Monitors progress
 - Recovers from problems
 - Tells you when your work is complete
- If there is scope for parallelism, it can also decompose your work into convenient execution units based on the available resources, data distribution

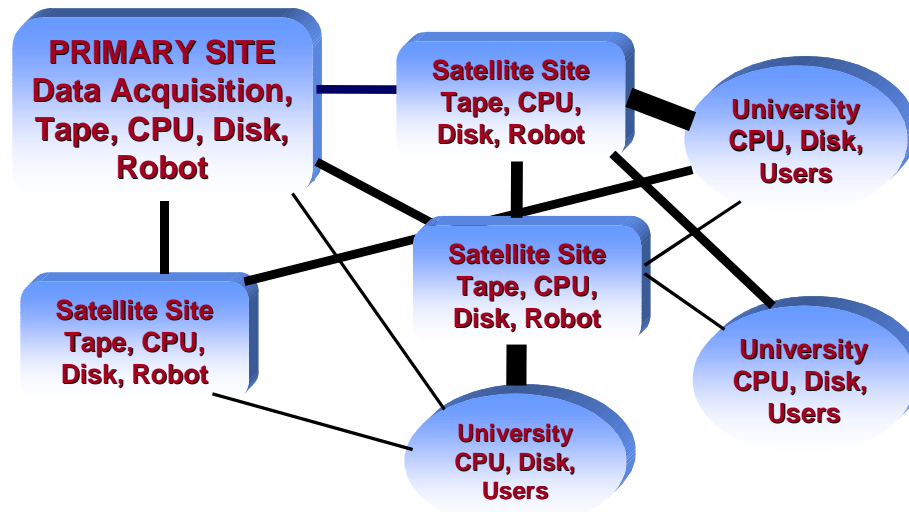


PPDG GRID R&D

Richard Mount, SLAC



PPDG Multi-site Cached File Access System



PPDG

November 15, 2000

LHC Computing Review



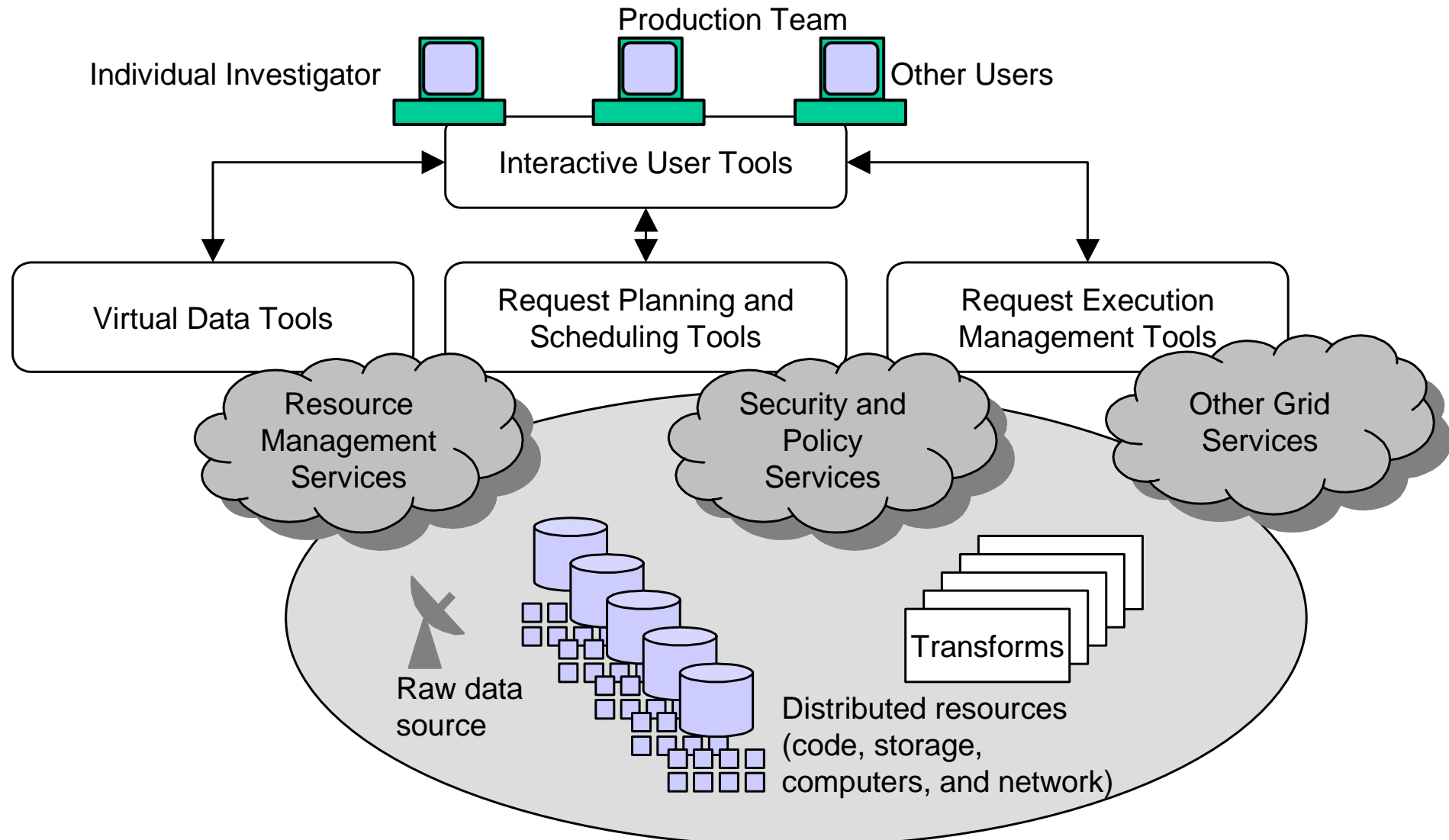
GriPhyN Overview

(www.griphyn.org)

- 5-year, \$12M NSF ITR proposal to realize the concept of virtual data, via:
 - 1) CS research on
 - Virtual data technologies (info models, management of virtual data software, etc.)
 - Request planning and scheduling (including policy representation and enforcement)
 - Task execution (including agent computing, fault management, etc.)
 - 2) Development of Virtual Data Toolkit (VDT)
 - 3) Applications: ATLAS, CMS, LIGO, SDSS
- PIs=Avery (Florida), Foster (Chicago)



User View of PVDG Architecture





GRID Computing

- GRID computing is a very hot topic at the moment.
- HENP is involved in many GRID R&D projects, with the next steps aimed at providing real tools and software to experiments.
- The problem is a large one and it is not yet clear that the concepts will be turned into effective computing.
 - CMS@HOME?



Conclusions

- Computing in HENP is a vital tool for extracting physics results, especially as the volume of data continues to dramatically increase.
- The challenges grow rapidly, certainly as fast or faster than advances in computing hardware and software techniques.
- Future experiments/accelerators/theory will continue to more heavily rely on computing, and new ideas like GRID computing will be pursued.